The Deepfake Universe Apocalypse?

Nadisha-Marie Aliman[1][0000−0003−3049−9327] and Leon Kester[2][0000−0002−8565−3902]

[1] Utrecht University, Heidelberglaan 8, 3584 CS Utrecht, The Netherlands
[2] TNO, Oude Waalsdorperweg 63, 2597 AK, The Hague, The Netherlands
leon.kester@tno.nl

Abstract. Could 2024 be the year heralding what one could term the *deepfake universe apocalypse* scenario or could it be the year that a future history of science may e.g. interpret as the year of the first literally universe-sized algorithmic hype bubble? This commentary introduces the metaphor of "GPT-Universe" and the assumptions hidden beneath it.

Keywords: Epistemology · Intelligence · Creativity · Consciousness.

1 The Problem: Hidden Assumptions

A part of the science community seems to believe in the probabilistic prophecy that there is a large chance that an algorithmic superintelligence able to surpass the entire present-day humanity in all tasks of interest to the whole of current humanity will be built by human civilization in a few thousand days (for clarity, we interpret this new time unit to signify a period spanning between ca. two thousand days and three thousand days). In a nutshell, a superintelligent *God*-like *alien* monster algorithmic *species* is expected to emerge within a few years. Historically speaking, among many others, a subset of the tasks of interest to present-day human civilization include scientific musings about a full planetary-scale, stellar-scale and possibly galaxy-scale civilization which is reflected in the breadth of topics analyzed in the decades-old search for extraterrestrial intelligence [8] (SETI) research[3]. In short, current proponents of algorithmic superintelligence immanency seem to assume that within a few thousand days, a humanity-made algorithm (a form of what we will refer to as "GPT-Universe") will *automate all of science* and thus be able to achieve all of the following successive unprecedented algorithmic miracles with an arbitrary lower latency than the entire human civilization ever could: 1) becoming a full planetary-scale entity with the ability to reach a power production matching terrestrial insolation magnitude, 2) becoming a stellar-scale entity, 3) becoming an intergalactic entity able to physically create a new universe [6, 13, 27] using a Planck collider [28, 22] and finally since the universe would then be algorithmic

[3] Why SETI scales offer a particularly suitable inspiration to identify multiple successive civilization-level tasks of interests required for a more rigorous scientific evaluation of algorithmic superintelligence claims has been described elsewhere [1, 2, 4].

and unexplainable allowing arbitrary algorithmic shortcuts to power production, nothing would keep this algorithm from 4) colonizing the entire universe (in the case of GPT-Universe, so as to achieve a non-sensical monologue with more and more copies of GPT-Universe instantiated on more and more matter) and eventually reproducing this catastrophe further by automating the creation of baby universes with similar fates. However, when faced with such extravagant predictions, it is important to transparently identify and highlight potential hidden core assumptions. Firstly, it is assumed that 1) *people* are entirely reducible to an algorithm. Secondly, it also presumes that 2) the *universe* is entirely reducible to an algorithm. The latter subtly follows from the assumption that *all* of science (i.e. including cosmology) can be automated – an idea which would entail the absolute algorithmiticity of the cosmos since it surmises that cosmological evolution could be reliably algorithmically predicted *a priori*. Before elaborating on the weakness of the mentioned assumptions in the light of modern insights from multiple disciplines ranging from physics over complexity science to biology, it is vital to note that for a person to select those premises implies the need to believe that life is and has always been an algorithmic simulation which naturally inherently signifies that the universe would be unexplainable and incomprehensible. It is a science-defeating position because it outstrips "science" of its meaning – by what people who self-label as algorithms and still claim to construct emotional states such as "fear of death through superintelligence" risk to loose their credibility because algorithmic life and consciousness represent an oxymoron [7]. In brief, philosophically speaking, there is not even a meaningful death for an entity without agency that never lived. Moreover, the denial exhibited by projecting the fear of a superintelligent algorithm to the *future* may need to be upended since a superintelligent algorithm would already have been the *past* cause of the algorithmic simulation within which those entities assume to appear – by what the scenario is inevitable *ab initio*. No company ceasing to build algorithms today could save believers from the algorithmic nature of the simulation they suppose to inhabit given the presupposed zombie-like state. No company claiming to be able to control an algorithm that would be superintelligent in relation to the entire human civilization could have any effect on the algorithmic state of the purported simulation. While the scenario of being an algorithmic simulation represents an interesting metaphysical reflection, it is not amenable to experimental problematization (e.g. the probability of doom can be set to 100 % today and be kept at that level forever since no amount of failures to build a superintelligent algorithm will ever be accepted to signal its impossibility, a problem similarly known from doomsday cults [40]) and does not belong to the realm of science due to the *supernatural* state of the implied external simulator separated from the universe. Moreover, the assumed lack of agency coupled to the idea of being caught in an algorithmic simulation scenario could transfer the genuine believer into a state slightly resembling a phenomenon in schizophrenia referred to as thought insertion (the delusion that one's thoughts are not one's own, but belong to someone else and have been inserted into one's mind). This perspective may not offer the most robust basis for security and safety measures.

2 A Solution: Deconstructing Hidden Assumptions

Firstly, as already collated in an earlier overview [3], more than a dozen impossibility statements stemming from diverse disciplines (see e.g. [4, 9, 14–16, 19–21, 23, 24, 29, 31, 32, 34, 36–38]) – ranging from i.a. physics over complexity science and philosophy of science to biology – provide strong reasons to reject[4] the first assumption (the statement that *humanity* would be reducible to an algorithm) – irrespective of the currently salient groupthink. Secondly, complementarily, many modern frameworks in physics provide robust new explanations that strongly reject the second assumption (the claim that *the universe* would be reducible to an algorithm). While one could state that there once was a certain tendency in physics in the past to attempt to achieve the formulation of a *final* absolute theory of everything after which humanity could retire having reached epistemic stasis in a reducible algorithmic cosmos, many physicists came up with very different conclusions in recent years. In the quantum gravity framework of Gomes et al. [15], the universe is expounded to be incompressible and irreducible whereby *"physical reality is not to be replicated in digital information, with the inevitable consequence that human-like AGI is equally impossible to attain in digital format"* [15]. Another example is Tim Palmer's concept of the universe as evolving on a *non*-algorithmic fractal invariant set [35]. Following the cosmologist Marina Côrtes, *"[...] the computer that would be able to simulate the entire set of unique events that happen in the universe would have to be the universe itself. We'd need another universe if we have the ambition to exhaustively express and reproduce the complete set of evolution laws that govern ours"* [10] which signifies that *"nature allows us to understand her, but she keeps a degree of novelty in store. Every now and then, she will surprise us with a combination of events that has as yet never happened [...]"* [10]. The physicist Lee Smolin is known for his concept of precedence [39] implying that the outcomes of interactions in nature are *not* cast in stone [10]. Overall, quantum physics describes a *participatory* universe [12, 26, 30] where the inherently non-algorithmic quantum measurements are an irreducible part of cosmological evolution. Following Thomas Hertog who authored a book [18] about Stephen Hawking's last theory, the act of observation in quantum mechanics coming from the environment itself caused early physical laws to evolve [41]. On the whole, because science (including cosmology) is a task of interest to current humanity, it is irrational to continue widely propagating the idea that an algorithmic superintelligence that will make humanity obsolete and after which there is epistemic stasis can be built by humanity or its algorithms. Strikingly, toward the end of his life, Stephen Hawking [17] remarked: *"Some people will be very disappointed if in the end there is no ultimate theory. I used to belong to that camp. I' m now glad that our search for understanding will never come to an end, and that we will always have the challenge of new discovery. Without it, we would stagnate."*

[4] How the topic could now be studied *scientifically* including amenability to experimental problematization in the deepfake era has been elucidated earlier [1, 2, 4, 5].

3 Conclusion

Certain entities believe that a superintelligent God-like alien monster algorithmic species is expected to emerge within a few years. In Section 2, we elucidated why the latter is paired with two core hidden assumptions that do *not* stand up to modern scientific scrutiny: 1) the assumption that *people* are reducible to an algorithm, 2) the assumption that *the cosmos* is reducible to an algorithm. If humanity would be part of an algorithmic simulation scenario, life and agency would be oxymorons while the denial exhibited by projecting the fear of the superintelligent algorithm to the *future* needs revision since this superintelligent algorithm would already have been the *past* cause of the algorithmic simulation within which those entities assume to appear – by what the scenario is both inescapable and inevitable *ab initio* and per definition. The latter bears the risk of inspiring a doomsday cult for self-proclaimed zombies fearing a deepfake death scenario, regretting a never-lived life. Having said that, we estimate that the main risk lies *not* in the doomsday metaphysical considerations per se which may be a stage of human development that may transfigure with time (e.g. via a coming scientific paradigm shift) but instead clearly in the currently escalating *misdirection* game [33] that certain commercial entities intentionally feeding on those fears play with humanity. Why advertising an *impossible*-to-build "potentially" Utopian or alternatively humanicidal universe-controlling algorithmic superintelligence if not for greedy profit and short-term attention mongering?

As stated by Jaron Lanier, *"[...] reality is irrepressible"* [25]. The cosmos is *not* reducible to an algorithm. As expressed by Stephen Hawking toward the end of his life, one cannot *"[...] view the universe from the outside"* [17]. It is illogical to keep claiming that the automation of all of science which includes *cosmology* is around the corner or can ever be achieved. As stated in another paper [5], *relative* to present-day humanity, an algorithmic superintelligence would be an algorithm able to generate arbitrary many successive *civilization*-level scientific paradigm shifts with arbitrary higher reliability and arbitrary lower latency than the entire present-day human civilization could. We maintain that it is impossible for present-day humanity or its algorithms to build this epistemic perpetuum mobile (see the discussion in Section 2). More generally, it is impossible [2] for any civilization and its algorithms to reliably build an algorithm that would genuinely be superintelligent in relation to that civilization[5]. What could the year 2024 mark in the history of science and technology? The year where people succeeded in co-constructing the up to now biggest possible delusion and epistemic self-sabotage (risking to cause *"limited, ineffective solutions"* [11]) serving the marketing of algorithms, the first literally *universe*-sized algorithmic hype bubble for which Gartner hype cycle heuristics could never have been prepared?

[5] For those who regard the definition of algorithmic superintelligence provided in [5] as too strict and prefer weaker definitions, there would then be no reason to label the less-capable entity as superintelligent in relation to the entire human civilization and to claim that the entity would be able *to make humanity as a whole obsolete* since there would always be scientifically analyzable tasks of interest left for human civilization that may even need better and better modes of enhancing collaboration.

References

1. Aliman, N.M.: Cyborgnetic Invariance. Aliman, Nadisha-Marie (2023)
2. Aliman, N.M.: Acentric Intelligence. PhilPapers (2024)
3. Aliman, N.M.: Epistemic Doom In The Deepfake Era. PhilPapers (2024)
4. Aliman, N.M.: Responsible AI Control. In: Trustworthiness and Responsibility in AI – Causality, Learning, and Verification. Schloss Dagstuhl–Leibniz-Zentrum für Informatik (2024)
5. Aliman, N.M., Kester, L.: Science in the Age of Algorithmic Chrysopoeia? PhilPapers (2024)
6. Ansoldi, S., Guendelman, E.I.: Child universes in the laboratory. arXiv preprint gr-qc/0611034 (2006)
7. Bennett, M.T., Welsh, S., Ciaunica, A.: Why Is Anything Conscious? arXiv preprint arXiv:2409.14545 (2024)
8. Charbonneau, R.: SETI, artificial intelligence, and existential projection. Physics Today **77**(2), 36–42 (2024)
9. Corazza, G.E.: Beyond the adjacent possible: On the irreducibility of human creativity to biology and physics. Possibility Studies & Society **1**(1-2), 37–45 (2023)
10. Cortês, M.: FREE WILL AND THE ARROW OF TIME. In: TIME AND SCIENCE: Volume 3: Physical Sciences and Cosmology, pp. 327–350. World Scientific (2023)
11. Cortês, M., Liddle, A.R., Emmanouilidis, C., Kelly, A.E., Matusow, K., Ragunathan, R., Suess, J.M., Tambouratzis, G., Zalewski, J., Bray, D.A.: AI Horizon Scanning, White Paper p3395, IEEE-SA. Part I: Areas of Attention. arXiv preprint arXiv:2410.01808 (2024)
12. Danielson, D.L., Satishchandran, G., Wald, R.M.: Killing horizons decohere quantum superpositions. Physical Review D **108**(2), 025007 (2023)
13. Farhi, E., Guth, A.H., Guven, J.: Is it possible to create a universe in the laboratory by quantum tunneling? Nuclear Physics B **339**(2), 417–490 (1990)
14. Fokas, A.S.: Can artificial intelligence reach human thought? PNAS nexus **2**(12), pgad409 (2023)
15. Gomes, V.G., Cortês, M., Liddle, A.R.: Higher dimensional energetic causal sets. arXiv preprint arXiv:2309.08694 (2023)
16. Harisch, K.: Why Artificial General Intelligence Is and Remains a Fiction. osif.io (2024)
17. Hertog, T.: On the origin of time . https://www.youtube.com/watch?v=fY3MbKNNG8o (2023), Youtube, The Royal Institution; accessed 01-November-2024
18. Hertog, T.: On the origin of time: Stephen Hawking's final theory. Bantam (2024)
19. Jaeger, J.: Artificial intelligence is algorithmic mimicry: why artificial "agents" are not (and won't be) proper agents. Neurons, Behavior, Data analysis, and Theory pp. 1–21 (2024)
20. Jaeger, J., Riedl, A., Djedovic, A., Vervaeke, J., Walsh, D.: Naturalizing Relevance Realization: Why agency and cognition are fundamentally not computational. Frontiers in psychology **15**, 1362658 (2024)
21. Kauffman, S., Roli, A.: The world is not a theorem. Entropy **23**(11), 1467 (2021)
22. Lacki, B.C.: SETI at Planck energy: When particle physicists become cosmic engineers. arXiv preprint arXiv:1503.01509 (2015)
23. LANDGREBE, J., SMITH, B.: Intelligence. And what computers still can't do. Cosmos+ Taxis **12** (2024)

24. Lanier, J.: There Is No A.I. https://www.newyorker.com/science/annals-of-artificial-intelligence/there-is-no-ai (2023), The New Yorker; accessed 01-November-2024
25. Lanier, J.: What My Musical Instruments Have Taught Me. https://www.newyorker.com/culture/the-weekend-essay/what-my-musical-instruments-have-taught-me (2023), The New Yorker; accessed 01-November-2024
26. Lewton, T.: Black Holes Will Eventually Destroy All Quantum States, Researchers Argue . https://www.quantamagazine.org/black-holes-will-destroy-all-quantum-states-researchers-argue-20230307/ (2023), Quanta Magazine; accessed 01-November-2024
27. Loeb, A.: Was Our Universe Created in a Laboratory? Scientific American. Oct **15**, 2021 (2021)
28. Loeb, A.: Interstellar Partnership on a Planck Collider . https://avi-loeb.medium.com/interstellar-partnership-on-a-planck-collider-1f1117404733 (2024), Medium; accessed 08-October-2024
29. MARTINELLI, E.: Complexity and Particularity: An Argument for the Impossibility of Artificial Intelligence. Cosmos+ Taxis **12** (2024)
30. Misner, C.W., Thorne, K.S., Zurek, W.H.: John Wheeler, relativity, and quantum information. Physics Today **62**(4), 40–46 (2009)
31. Mueller, M.: The Myth of AGI. Internet Governance Project (2024)
32. Muraleedharan, A.: Turing Machines cannot simulate the human mind. arXiv preprint arXiv:2207.05700 (2022)
33. Nagy, P., Neff, G.: Conjuring algorithms: Understanding the tech industry as stage magicians. new media & society **26**(9), 4938–4954 (2024)
34. Palmer, T.: Human creativity and consciousness: Unintended consequences of the Brain's extraordinary energy efficiency? Entropy **22**(3), 281 (2020)
35. Palmer, T.N.: Undecidability, Fractal Geometry and the Unity of Physics. Undecidability, Uncomputability, and Unpredictability pp. 81–95 (2021)
36. Penrose, R.: Is Conscious Awareness Consistent with Space-Time Descriptions? In: Philosophy, Mathematics and Modern Physics: A Dialogue, pp. 34–47. Springer (1994)
37. Popper, K.R., Lindahl, B.I.B., Århem, P.: A discussion of the mind-brain problem. Theoretical Medicine **14**, 167–180 (1993)
38. Roli, A., Jaeger, J., Kauffman, S.A.: How organisms come to know the world: Fundamental limits on artificial general intelligence. Frontiers in Ecology and Evolution **9**, 806283 (2022)
39. Smolin, L.: Precedence and freedom in quantum physics. arXiv preprint arXiv:1205.3707 (2012)
40. Snow, R.L.: Deadly cults: The crimes of true believers. Bloomsbury Publishing USA (2003)
41. Turner, B.: 'Physics itself disappears': How theoretical physicist Thomas Hertog helped Stephen Hawking produce his final, most radical theory of everything . https://www.livescience.com/space/cosmology/physics-itself-disappears-how-theoretical-physicist-thomas-hertog-helped-stephen-hawking-produce-his-final-most-radical-theory-of-everything (2024), LiveScience; accessed 01-November-2024