

Revisiting Freedom at the 50th St. Gallen Symposium

Global Essay Competition

Innovators' Freedom to Challenge Our Paradigms: Why the Copernican Legacy Should Guide Our Progression through the Incipient Age of Artificial Intelligence

Author & Institution	Janik Festerling, University of Oxford
Short Essay Title	Innovators' Freedom to Challenge Our Paradigms
Full Essay Title	Innovators' Freedom to Challenge Our Paradigms: Why the Copernican Legacy Should Guide Our Progression through the Incipient Age of Artificial Intelligence
Essay Position	In defence of freedom
Word count	2098 (excl. abstract, keywords, endnotes, references)

Abstract

Despite originating from the innovative, yet niche work of an astronomically interested polymath living in the gloomy aftermath of the European medieval period, the *Copernican Revolution* is considered to be the historical progenitor of many freedoms that we enjoy in modern secularised societies. But while the notion of the Copernican Revolution primarily refers to its role as the igniter of a singular chain reaction with historical momentum, I argue that Copernicus' true *legacy* stands for a *freedom* in itself that has not lost its significance ever since its birth in the dead of the 16th-century: innovators' freedom to challenge the prevailing paradigms of their times. This essay is meant to be a thought-provoking treatise in defence of this freedom. Firstly, based on a brief historical excursus I derive the argument why innovators' influence on societal paradigms has always been a main driver of societal progression. Secondly, I consider why innovations related to the recent advancements in Artificial Intelligence (AI) may be the portents of another paradigm shift. Thirdly, I discuss why and how innovators' freedom to challenge the prevailing paradigms of *our* time is worth defending.

Keywords

Artificial intelligence, history of innovation, paradigm shifts, intergenerational change

When the European polymath Nicolas Copernicus (1473 – 1543) published his innovative but nevertheless niche work on how to compute the motions of stars and planets based on an alternative reference point – namely the sun instead of the earth – he probably had no idea that depriving the earth of its pivotal role in astronomical theory would ignite a historical chain reaction.ⁱ But generations born in the aftermath of Copernicus’ era gradually developed a completely different understanding of what they saw when they were stargazing and what the nature of their own creaturely existence might be within this stunning darkness they faced. Eventually, this slight perspectival shift, which later became known as the *Copernican Revolution*, translated into a major *paradigmatic* shift by heralding the decline of the church’s societal primacy, and therefore paving the way for the post-medieval development of society [1].

So, despite originating from the innovative, yet niche work of an astronomically interested polymath living in the gloomy aftermath of the European medieval period, Copernicus arguably remains the progenitor of many freedoms that we enjoy in modern secularised societies [1]. But while the notion of the Copernican Revolution primarily refers to his role as the igniter of a singular chain reaction with historical momentum, I argue that Copernicus’ true *legacy* stands for a *freedom* in itself that has not lost its significance ever since its birth in the dead of the 16th-century: innovators’ freedom to challenge the prevailing paradigms of their times.

This essay constitutes a thought-provoking treatise in defence of this freedom. Firstly, based on a brief historical excursus I derive the argument why innovators’ influence on societal paradigms has always been a main driver of societal progression. Secondly, I consider why innovations related to the recent advancements in Artificial Intelligence (AI) may be the portents of another paradigm shift. Thirdly, I discuss why innovators’ freedom to challenge the prevailing paradigms of *our* time is worth defending.

1. Human History: A By-Product of Human Innovation

Certainly, the history of human innovation is older than the Copernican Revolution, and there are several instances predating Copernicus’ work which have yet predetermined modernity with similar momentum through their paradigm shifts. For instance, the mechanical clock went far beyond the simple purpose of timekeeping by conveying the idea that there must be an objectively measurable truth behind the opaqueness of natural phenomena, therefore paving the way for the development of experimental methods and a major scientific paradigm shift [2]. Similarly, the perfection of the printing press and papermaking in the 15th-century resulted in the printing, distribution, and archiving of diverse knowledge at increasing scales and reduced the reliance on

verbal narratives or singular hand-written manuscripts [2]. But, even more importantly, it also ignited the gradual spread of literacy within society [3][4]. And then, very recently – at least from the perspective of a historian – an array of technological and scientific innovations finally came to birth as part of the 19th-century Industrial Revolution [2], which, among other things, successively modernised the traditional societal function of the family and the individual roles of its members [5][6].

So, to draw an interim conclusion, the societal depth and breadth of technological or scientific innovation has never been limited to its niche, that is, its immediate purposes and applications. Instead, innovators have always predetermined society's DNA, including our daily practices, value systems, beliefs, hopes, fears, or pleasures. As Shulevitz eloquently summarises: “[h]uman history is a by-product of human inventions” [7, para.8]. Therefore, our current situation in which we find ourselves in the first half of the 21st-century – and which remains arguably the most progressive state of humankind up to this point [8] – can be expressed as a function of those innovations that have paved the way.

But in light of this provocative assertion, another issue must be addressed: of course, there are serious risks that innovation naturally entails and that have become manifest many times in history whenever an invention – intentionally or unintentionally – translated into something that now seems backward or questionable in retrospect. For example, innovations in navigation and seafaring arguably heralded the age of European imperialism and colonialism [9], the military applications of nuclear fission introduced an ongoing era of mass destruction and nuclear deterrence [10][11], and, most importantly, the Industrial Revolution ignited those polluting and exploitative forces that contributed to the climate and sustainability crisis we face today [12].

But the question of main importance remains: do the implicit risks of a freedom justify its oppression? If human history had been left to the supporters and sympathisers of this view, we would not have been where we are today. Surely, while being kept in a state of prehistoric stagnation, we would have avoided many disastrous chapters of history, such as imperialism, colonialism, the threat of nuclear warfare and, after all, climate change. But if we assume that progression is the purpose of human endeavour, that is, the purpose of our existence and the meaning of our lives, then we must also accept its risks to reasonable extents. So, the reader may either curse or appreciate the current state of humanity and how we got here, but as we speak, the incipient age of AI gives rise to an array of new innovations, and we have to consider to what extent we grant AI-innovators the freedom to challenge the prevailing paradigms of our time.

2. Incipient Age of AI: Why Intelligent Machines Constitute a Paradigm Shift

We find ourselves in the middle what has been famously labelled as the *Fourth Industrial Revolution*, in which humanity undergoes the large-scale digitisation of society as a whole [13]. As part of this process, the increasing availability of digital data and computational power has resulted in rampant advancements in AI, which can be perceived as foretastes of humanity's potential to create machines that are able to compete with the once uniquely domains of human intelligence [14]. For instance, recent breakthroughs have exemplified how self-learning machines can outperform human world champions at the most complex board games, such as *chess*, *shogi* or *Go* – and notably without relying on any prior human wisdom regarding rules or strategies [15]. However, playing board games may certainly sound like a narrow niche for innovation, just like many of the other highly specialised applications in which AI already optimises our daily lives, such as recommendation engines, navigation, mobile-device operating systems, financial services, social media, or voice assistants. So, what's the last time such innovative but nevertheless niche breakthroughs turned out be epochal for human history? Well ...

At its very core, the pursuit behind AI-innovation has always been more than the optimisation of existing solutions, more than the mastering of specific tasks, more than a means to an end. Since the emergence of AI in the 1950s, mathematicians, computer scientists, engineers, biologists, philosophers and others have perceived themselves as “creators of life“ [16, p.237]. Hence, it is not surprising that most of today's world leading AI-innovators have a background in neuroscience or cognitive psychology [17], because, after all, the idea to make machines intelligent by programming self-learning artificial neural networks is inspired by computer scientists' belief that the human brain might learn in the same way [18], and that its perfect artificial reproduction must be possible, at least in theory [19].ⁱⁱ

Hence, for a long time AI-innovators may have been experts working in the niches of their respective disciplines, just like Copernicus working in his astronomical niche more than 500 years ago. But at this point, it becomes apparent that the impacts of their innovations will go way beyond the immediate purposes and applications, because creating machines that emulate or even surpass our intelligence may not only affect how we think about these machines, but also how we think about ourselves as humans! Of course, paradigm shifts are slow by their very nature and spread across several generations, but the depth and breadth of their impacts should not be underestimated [20]. For this reason, the societal discourse on AI-issues will probably change significantly once today's youngest generations come into power. This includes their relationship with machines,

their perspectives and trad-off choices related to privacy, or their ethical beliefs in the good and the bad – and it would probably feel very discomfoting for those who are in power today, but, naturally, they will be gone by then.

3. AI-Innovation & Progression: Why We Need to Defend the Freedom

As AI progresses, innovators' freedom to challenge our paradigms is at stake due to the increasing struggle to understand their work, the denial of their powerful breakthroughs, or the demonization of their pursuits as a whole. In other words, we face the threat of human nature to be apologetic, or as Wolfram points out: "it's another part of the Copernican story: [w]e used to think Earth was the center of the universe. Now we think we're special because we have intelligence, and nothing else does" [20, p.284]. Indeed, from a scientific perspective, there is an ever-shrinking reason to believe in our sacrosanct speciality: today's most plausible neuroscientific explanation for the underlying mechanics of our cognitive intelligence is that our brains perform computations when processing the complex data they receive from our senses in order to derive inferences and make predictions, just like state-of-the art machines in AI [22].ⁱⁱⁱ For this reason, the perspective emerges that machines will eventually surpass all innate limitations of human intelligence and might introduce a new scientific era in which our understanding of the world could be leveraged with *superhuman* insights revealed by AI (e.g. [23][24][25]).

But the need for an AI-driven societal paradigm shift goes far beyond the traditional fascination of scientists to augment their understanding of the world or simply push boundaries. In light of the contemporary challenges that society faces in the middle of its digitisation, we may realise that once again we must use the fruits of our progression to address challenges arising along the way. For instance, the rise of social media has brought out the most reprehensible, yet enduring facets of human behaviour like no other innovation before, such as the spread of disinformation, hate speech, violence, or child abuse [26][27]. In this context, the confrontations between Facebook founder and CEO Mark Zuckerberg and various politicians at his hearings illustrate the paradigmatic contrast: on the one hand, Mark Zuckerberg representing the innovator who seeks an AI-driven paradigm shift by emphasising how social media utterly relies on the proactive means of AI in order to fight these challenges and maintain a good purpose (e.g. [28]); on the other hand, various confused and helpless politicians of another generation struggling to grasp what he said [29]. Similarly, in a recent seminal paper Vinuesa et al. have pointed out how AI-applications could significantly leverage the achievement of humanity's *Sustainable Development Goals* by enabling innovation in various contexts [30].

Nevertheless, a plain, yet critical question remains: do we really need to *defend* innovators' freedom to challenge our paradigms? After all, the key drivers of AI-innovation, such as Amazon, Alphabet, Apple, Facebook, or Microsoft, remain the most powerful organisations of our time, and in terms of governmental regulations and control they have barely faced policies that would limit their freedom to change society – in fact, they did already! Subsequently, given all the obvious risks of AI-innovation that are currently discussed, is it really worth defending this freedom instead of urging for its regulation and control?

I conclude on three reasons why our final answer should be affirmative. Firstly, defending a freedom and considering its ethical boundaries or regulatory control is neither mutually exclusive nor mutually harmful. In fact, we would probably agree that an important achievement of human history has been the development of ethics and regulations guiding innovation – and even the key drivers themselves begin to urge for *concrete* regulations around AI [31][32].

Secondly, whether a freedom is worth defending does not only depend on the criticality of its integrity in society, but also how important it is for society's overall progression. So, given that we have seen how closely human progression is associated with the history of innovation due to its paradigmatic powers, its defence should be worthwhile at all times.

Thirdly, and most importantly, we should keep in mind that defending a freedom is not only a matter of defending the ones who exercise it, but it is also about defending the entirety of its beneficiaries. In particular, paradigm shifts are, and always have been, placed in the hands of younger generations who are born in the middle or in the aftermath of the revolution itself. When Nicolas Copernicus succumbed to an illness in 1543, modernity was still in its infancy. But he left something that enabled following generations to embrace his legacy and translate it into human progression. It was not only his freedom, it was also theirs, and, after all, ours!

Endnotes

ⁱ Copernicus' proposition, originally embedded in a complex assembly of graphs and formulas, that our designated place of earthly existence might only be one out of many astronomical objects in the universe constituted a major opposition to the church's religious narrative in which earth and its creatures remained the creations of the Creator in the centre of the universe [1].

ⁱⁱ On philosophical grounds, this so-called *functionalist* view challenges René Descartes' dualism on the distinct nature of the conscious mind and unconscious matter, in which Descartes concludes that the theoretical question whether machines can have mindful brains – or differently spoken, the question whether the inner mechanics of non-living entities in the physical world can possess mentality – must be negated [33]. In contrast, modern opponents of functionalism, like Hubert Dreyfus [34][35] or John Searle [36][37], argue that even the most accurate and sophisticated computational model of the human brain could not be considered as a thinking and conscious mind in itself. Therefore, the debate comes down to the ancient Cartesian question whether the existence of thought and mind *does* or *does not* depend on the existence of an 'I', and the functionalists' answer remains no, while dualist opponents continue to insist on yes [16].

ⁱⁱⁱ In fact, even at this point the remaining difference of major importance might be that from a conceptual perspective the natural learning of the human brain can combine what has been traditionally separated within AI-innovation, namely the *connectionist* approach based on artificial neural networks and the *symbolic* approach based on generative Bayesian models [22][38][39]. Although the recent breakthroughs in AI have already proven that machines can outperform humans within each approach (e.g. the connectionist approach has enabled deep-learning algorithms to develop board game strategies that were unknown for thousands of years [40]), their combination within a single computational entity remains very promising (e.g. [41]).

References

- [1] T. S. Kuhn, *The Copernican Revolution: Planetary Astronomy in the Development of Western Thought*, Reprint. London, England, UK: Harvard University Press, 2003, 1957.
- [2] L. Mumford, *Technics and Civilization*. London, England, UK: University of Chicago Press, 2010, 1934.
- [3] W. J. Ong, *Orality and Literacy: The Technologizing of the Word*. London, England, UK: Routledge, 1988.
- [4] E. L. Eisenstein, *The Printing Press as an Agent of Change : Communications and Cultural Transformations in Early Modern Europe*. Cambridge, England, UK: Cambridge University Press, 1980.
- [5] W. F. Ogburn and M. F. Nimkoff, *Technology and the Changing Family*. Boston, MA, US: Houghton Mifflin, 1955.
- [6] R. F. Winch, *The Modern Family*, 3rd ed. New York, NY, US: Holt, Rinehart & Winston, 1971.
- [7] J. Shulevitz, “Alexa, How Will You Change Us?,” *The Atlantic*, Washington DC, US, Nov-2018.
- [8] S. Pinker, *Enlightenment Now: The Case for Reason, Science, Humanism and Progress*. London, England, UK: Penguin Books, 2019.
- [9] N. Canny and P. Morgan, “The Making and Unmaking of an Atlantic World,” in *The Oxford Handbook of the Atlantic World: 1450-1850*, N. Canny and P. Morgan, Eds. Oxford, England, UK: Oxford University Press, 2011, pp. 1–18.
- [10] K. N. Waltz, *Theory of International Politics*. New York, NY, US: Random House, 1979.
- [11] K. N. Waltz, “The Spread of Nuclear Weapons: More May Be Better: Introduction,” *Adelphi Pap.*, vol. 21, no. 171, p. 1, 1981.
- [12] S. Solomon, *Climate Change 2007 : The Physical Science Basis*. Cambridge, England, UK: Published for the Intergovernmental Panel on Climate Change [by] Cambridge University Press, 2007.
- [13] K. Schwab, *The Fourth Industrial Revolution*. Geneva, Switzerland: World Economic

- Forum, 2017.
- [14] R. Kurzweil, *The Age of Intelligent Machines*. Cambridge, MA, US: MIT Press, 1990.
 - [15] D. Silver *et al.*, “A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play,” *Science* (80-.), vol. 362, no. 6419, pp. 1140–1144, 2018.
 - [16] S. Turkle, *The Second Self: Computers and the Human Spirit*, Twentieth. London, England, UK: The MIT Press, 2005.
 - [17] V. Ramakrishnan, “Will Computers Become Our Overlords?,” in *Possible Minds: Twenty-Five Ways of Looking at AI*, J. Brockman, Ed. New York, NY, US: Penguin Press, 2019, pp. 181–191.
 - [18] N. Thompson, “An AI Pioneer Explains The Evolution of Neural Networks,” *WIRED*, 2019. [Online]. Available: https://www.wired.com/story/ai-pioneer-explains-evolution-neural-networks/?utm_campaign=the_algorithm.unpaid.engagement&utm_source=hs_email&utm_medium=email&utm_content=72659496&_hsenc=p2ANqtz-8bs2iv7YTA5GAstR1JvuB4Qond4xhyEHoB6UP_21cUs_XWPwttnrXu1WCGM6TC. [Accessed: 18-May-2019].
 - [19] F. Wilczek, “The Unity of Intelligence,” in *Possible Minds: Twenty-Five Ways of Looking at AI*, J. Brockman, Ed. New York, NY, US: Penguin Press, 2019, pp. 66–75.
 - [20] M. Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence*. London, England, UK: Penguin Books, 2018.
 - [21] S. Wolfram, “Artificial Intelligence and the Future of Civilization,” in *Possible Minds: Twenty-Five Ways of Looking at AI*, J. Brockman, Ed. New York, NY, US: Penguin Press, 2019, pp. 266–284.
 - [22] A. Gopnik, “AI Versus Four-Year-Olds,” in *Possible Minds: Twenty-Five Ways of Looking at AI*, J. Brockman, Ed. New York, NY, US: Penguin Press, 2019, pp. 219–230.
 - [23] M. Kosinski, “Computational Psychology,” in *Advanced Social Psychology: The State of the Science*, 2nd ed., R. F. Baumeister and E. J. Finkel, Eds. Oxford, England, UK: Oxford University Press, 2019, pp. 499–523.
 - [24] S. Strogatz, “One Giant Step for a Chess-Playing Machine,” *The New York Times*, vol. 221, no. 3, New York, NY, US, p. D1, Jan-2019.

- [25] N. Bostrom, *Superintelligence: paths, dangers, strategies*. Oxford, England, UK: Oxford University Press, 2014.
- [26] C. Gray, “As a Facebook moderator I saw the worst of humanity. We need to be valued,” *The Guardian*, 2019. [Online]. Available: <https://www.theguardian.com/commentisfree/2019/dec/09/facebook-moderator-worst-of-humanity-valued>. [Accessed: 23-Jan-2020].
- [27] C. Newton, “The Trauma Floor: The Secret Lives of Facebook Moderators in America,” *The Verge*, 2019. [Online]. Available: <https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona>. [Accessed: 23-Jan-2020].
- [28] Facebook, “An Update on How We Are Doing At Enforcing Our Community Standards,” *Community Standards Enforcement Report*, 2019. .
- [29] BGOV, “Transcript of Mark Zuckerberg’s Senate hearing,” *The Washington Post*, 2018. [Online]. Available: <https://www.washingtonpost.com/news/the-switch/wp/2018/04/10/transcript-of-mark-zuckerbergs-senate-hearing/https://www.washingtonpost.com/news/the-switch/wp/2018/04/10/transcript-of-mark-zuckerbergs-senate-hearing/>. [Accessed: 23-Jan-2020].
- [30] R. Vinuesa *et al.*, “The role of artificial intelligence in achieving the Sustainable Development Goals,” *Nat. Commun.*, vol. 11, no. 1, p. 233, 2020.
- [31] S. Pichai, “Why Google thinks we need to regulate AI,” *Financial Times*, 2020. [Online]. Available: <https://www.ft.com/content/3467659a-386d-11ea-ac3c-f68c10993b04>. [Accessed: 24-Jan-2020].
- [32] H. Murphy, “Zuckerberg calls for more regulation of big tech,” *Financial Times*, 2019. .
- [33] M. Wheeler, “God’s Machines: Descartes on the Mechanization of Mind,” in *The Mechanical Mind in History*, P. Husbands, O. Holland, and M. Wheeler, Eds. Cambridge, MA, US: Oxford University Press, 2008, pp. 307–330.
- [34] H. L. Dreyfus, *What computers can’t do: A critique of artificial reason*. New York, N.Y., US: Harper & Row, 1972.
- [35] H. L. Dreyfus, *What computers still can’t do: A critique of artificial reason*, 3rd ed. London, England, UK: MIT Press, 1992.

- [36] J. R. Searle, “Chinese room argument,” *Scholarpedia*, vol. 4, no. 8, p. 3100, 2009.
- [37] J. R. Searle, “Is the Brain’s Mind a Computer Program,” *Sci. Am.*, vol. 262, no. 1, pp. 26–31, 1990.
- [38] J. Pearl, “The Limitations of Opaque Learning Machines,” in *Possible Minds: Twenty-Five Ways of Looking at AI*, J. Brockman, Ed. New York, NY, US: Penguin Press, 2019, pp. 13–19.
- [39] R. Velik and D. Bruckner, “Neuro-Symbolic Networks: Introduction to a New Information Processing Principle,” in *IEEE 2008 - 6th IEEE International Conference on Industrial Informatics*, 2008, pp. 1042–1047.
- [40] C. Anderson, “Gradient Descent,” in *Possible Minds: Twenty-Five Ways of Looking at AI*, J. Brockman, Ed. New York, NY, US: Penguin Press, 2019, pp. 143–150.
- [41] J. Mao, C. Gan, P. Kohli, J. B. Tenenbaum, and J. Wu, “The Neuro-Symbolic Concept Learner: Interpreting Scenes, Words, And Sentences From Natural Supervision,” in *ICLR 2019 - Seventh International Conference on Learning Representations*, 2019.