

An Evaluation of the Potential Future for AI and the UN SDGs

(Full assignment question in appendix 1)

Student Name: Harry Hare

Institution: Durham University Business School

Level 3 Undergraduate

Category one

**Declaration: This essay is entirely my own work, and the use of generative AI
has not been used in the writing process.**

Abstract

‘Humanity is at a crossroads’ (Poole, 2024, p.XI). The opening words to the preface of Eve Poole’s *Robot Souls* perfectly summarise humanity’s AI predicament; on the one hand, AI offers countless opportunities for human progression, enabling previously impossible levels of research, development and creativity. However, if the current AI trajectory continues, one of exploitation, surveillance and ‘technofeudalism’, we risk a ‘civilizational crisis that humanity has never encountered or navigated before’ (Vallor, 2024, p.10). The direction humanity takes regarding AI must, therefore, depart from the corporate-dominated neoliberalist paradigm and instead adopt the virtues and ethics of Kant and Aristotle, with a collectivised goal of using ‘AI for good’ in businesses, business schools and governments alike. Using the UN’s 17 sustainable development goals as a template (UN, 2024), this essay will summarise both the need for and challenges of deploying AI for good before evaluating the threat AI poses and how these threats can be averted.

The AI conundrum: using machine learning for good in the face of adversity

In the opinion of Vallor (2024), the threat of AI is not external but comes from within humanity. Drawing on the story of Ovid’s Narcissus, the beautiful man whose obsession with his reflection led to his death, she develops the metaphor of AI as a mirror. This metaphor serves a valuable purpose, illustrating how the nature of machine learning is what makes it so dangerous; AI mirrors project back to us a vision of human intelligence ‘constructed entirely from the amalgamated data of humanity’s past’ (ibid, p.6), freezing us in a place of constant fascination of what has been. However, AI’s reflective nature is not equal or reasonable; instead it mirrors morals and virtues of the post-industrial societies that built them (ibid). Most AI

systems are developed in the Global North, neoliberal Western arenas where unbridled freedom, deregulation and profit-seeking aim to undermine and attack any notion of a 'society' (Brown, 2018). Assets that humanity once championed, creative thought, moral ambition and wisdom, have been drowned out by the 'AI mirror', leaving us with systems that directly oppose and suffocate traditional human virtues. The results of this faux-reflection, created by a small and elite subsection of society, are AI mirrors that do not represent the needs and goals of humanity but instead enforce historical patterns of oppression, hierarchical control and exclusion.

Vallor argues that to 'reclaim our humanity in an age of machine thinking', there is little use in realigning AI with contemporary human values, as these values led us here in the first place. Instead, a revaluation of humans' fundamental relationship with each other and technology needs to be adopted. Building on Kantian ethics (Kant, 1997), Vallor suggests moral and intellectual virtues, including wisdom, courage, justice and honesty, must be maintained and promoted when developing and using AI. Furthermore, Vallor (2016, p.19) builds on the Aristotelian principle that 'the virtuous person not only tends to think and act rightly but also to feel and want rightly' and argues that for a sustainable AI future, these same virtues must be integrated into humanities relationship with technology. In Vallor's opinion, implementing these virtues is vital in the continuous and urgent struggle against the existing hierarchical power dynamics the current age of machine thinking is enforcing.

If the virtues and morals Vallor suggests are implemented, AI has potential to overhaul how humanity develops sustainably. AI for Good (UN, 2024) is a UN-led initiative aiming to provide AI-driven solutions to problems aligned with the UN's 17 SDGs. It aims not only to find trustworthy AI applications but also to advance AI

governance for sustainable development. A survey of experts carried out by McKinsey (2024) found that AI can specifically positively impact five of the SDGs: Good Health and Well-Being, Quality Education, Climate Action, Affordable and Clean Energy and Sustainable Cities and Communities. AI's ability to aid in tasks such as providing medical diagnoses and educational resources to remote or excluded communities, creating algorithms to examine climate data and building or controlling sustainable and clean energy solutions has the potential to be revolutionary for humanity; never before has data collection, analysis and distribution been so easy and accessible. If applied to the 'restorative care and services of life', AI opens new doors to previously thought impossible levels of equality and development, and, if implemented adequately alongside the 17 SDGs, it could play a crucial role in delivering humanity from looming environmental, political and social crises.

However, there is an arduous yet nonetheless imperative task in defining what using 'AI for good' means. Kant (1997) suggested doing no harm was relatively easy; actions such as stealing and lying are always wrong and are therefore defined as perfect duties. Doing right (or good), on the other hand, imperfect duties, is open-ended; it is up to the discretion of the actor, which in the case of AI, and according to Vallor, currently resides with 'Big Tech' and neoliberalist focused governments. She argues that most current 'AI for Good' proposals are still enthralled by the AI mirror, their own Narcissuses, reflecting the same functions of AI that have caused harm elsewhere (Vallor, 2024). AI's reflection of the past poses a challenge to corporate hopes of benefitting from AI systems; currently, even when adopting an altruistic approach, they are still forced into old habitual practices, surveillance and predicting and classifying one another (ibid). In Stiglitz's (2024,

p.290) view, these practices define a neo-authoritarianism, 'made worse by advances in science and technology'. There is a case for utilising this neo-authoritarianism by implementing AI's surveillance and persuasion potential to enforce rather than suggest a new paradigm of sustainable thinking. According to Li & Shapiro (2020), this argument holds particular resonance at a time when liberal democratic solutions to environmental problems routinely fail. However, in practice, examples such as China's 'environmental authoritarianism' tend to prove that draconian measures undertaken in the guise of green thinking lead to further removal of human liberties, preserving power and social control for those in charge. Furthermore, and crucially, AI has a fundamental problem with opacity. AI's very nature makes it enigmatic; because AI essentially develops its own model, it can quickly morph into a system impossible for most humans to scrutinise or comprehend (Stahl, Schroeder, Rodrigues, 2023). This creates a stark problem for businesses and governments alike, who run the risk of being faced with AI problems, or indeed solutions, they do not have the understanding to scrutinise or critique. The complexity of AI models means they can 'manufacture plausible facsimiles of moral reasoning' (Vallor, p.111) that may go unnoticed by the majority and, therefore, be implemented and utilised in various walks of political, social and business lives. The implications of this overconfidence in AI's proficiency can have an extensive range of repercussions, from relatively mundane issues of students getting incorrect answers from their ChatGPT assignment preparation to the catastrophic, like in the case of UnitedHealth's faulty AI model nHPredict, which denied elderly patients claims for extended care, resulting in 90% of rejected claims having to be overturned manually (Laney, 2023).

Suppose the current mode of AI usage is defined by reflections of a neoliberal, market-based and profit-orientated past, then there is a case to be made for an overhauling of the very structures of our businesses and governance. If the radical approach of environmental authoritarianism cannot function justly within the current socioeconomic climate without perpetuating exploitation, a responsibility lies with institutions to reject the neoliberal market-based paradigm of profit maximisation and instead adopt a justice failure approach. The justice failure approach directly counters Friedman's interpretation of social responsibility; it recentres corporate objectives away from profit and instead insists on companies to incorporate and develop solutions to social, environmental and development problems. In the past, models such as Elkington's triple bottom line, which examines a company's social, environmental and economic impact, have been suggested, providing a positive framework to move towards a justice failure approach. However, Elkington (2018) admits his TBL has ultimately failed, as early aims of 'provoking deeper thinking about capitalism' have been dismantled and replaced with a trade-off mentality. Instead, the corporate approach to AI must aim for a sustainability equilibrium, where the creation of social value balances the relationship between economic goals of profit and shareholder returns. Raworth (2018) advocates for a doughnut economy model, whereby all institutions reside where humanity's basic needs are provided for while avoiding further exploitation of the climate, society, and natural resources. By adopting these radical new ways of approaching corporate responsibility and organisation and stimulating the systemic change of humanities very virtues and morals that Vallor suggests, there is hope for a future of using AI for good.

The dangers of AI and a new paradigm

The issues with AI in its current iteration are, however, unfortunately not merely limited to holding back positive advancements or reinforcing past hierarchies of injustice or unsustainability. Instead, AI poses a very real and active threat, exaggerating an increasing trend towards a post-capitalist mode in which the majority are feudalised by a minority few: the ‘technofeudal lords’ of Big Tech and their respective computing systems. Streek (2016) suggests the era of traditional capitalism is ending as growth slows and the economy stagnates, creating a desperate struggle for exploitation and accumulation as capitalists desperately cling to the hope of ever-higher profits. Varoufakis (2023), however, argues that the very notions of traditional ‘capital’ and ‘capitalism’ have been mutated into something entirely novel: an era of ‘cloud capital’ whereby the two traditional pillars of markets and profits have been ‘evicted from the epicentre of our economic and social systems’ (ibid, p.8). In this new technological era, Streek’s desperate, profit-seeking capitalists are replaced with something far more powerful and dangerous, ‘technofeudal’ lords who control and charge rent for the platforms that make up almost every aspect of modern life. Companies like Alphabet, Meta and Microsoft now only superficially sell us physical commodities; cheap phones, home devices, and other gadgets are merely façades for their real aims. Instead, these companies have turned their users into products; they capture the attention of unbeknownst participants with seemingly harmless applications, such as TikTok or Facebook, before harvesting their data and selling it on to other Cloud Capitalists to be used for unprecedented levels of intrusive marketing and control. As Varoufakis (2023, p.52) summarised, the new-age cloudalist is no longer interested in competing within a capitalist market, but instead in getting ‘us to exit capitalist markets altogether’.

The threat of this new ‘technofeudal’ system is only inflated by the same opacity problem of AI that makes using these systems for good so arduous. The AI systems we knowingly, or often unknowingly, interact with daily are presented as forces for good. Meta’s ‘Metaverse’ is marketed as a platform for ‘new ways to connect and share experiences’ (Meta, 2024), Open AI’s claims to ‘ensure that artificial general intelligence benefits all of humanity’ (OpenAI, 2024) and Apple’s ‘Apple Intelligence’ is simply presented as ‘AI for the rest of us’ (Apple, 2024). The inclusive presentation of these systems makes them all the more dangerous, ‘unlike the tyrants defeated by past movements of democratic resistance, these tools are presented to us ... not as selfish oppressors, but as benign and neutral agents of our will’ (Vallor, p.131). This feigned benignity has led to the systems of injustice that Vallor warned about not only being mirrored but also amplified. Issues of privacy, surveillance and manipulation have increasingly moved beyond the physical world and onto the cloud, a place where regulation and governance are limited and power lies with corporations rather than traditional governmental institutions, leaving room for malpractice. Incidents of this have become increasingly commonplace; for example, Meta’s Cambridge Analytica scandal, where upwards of 87 million Facebook users’ data was shared with a private consulting firm without consent and allegedly used in voter targeting campaigns for Donald Trump’s 2016 presidential campaign (McCallum, 2022). In a similar vein, the UK government has suggested algorithmic clustering of far-right individuals on X was partially to blame for the anti-immigration riots seen in the summer of 2024, as people were fed misinformation online, leading to violent and devastating real-world consequences (UK Parliament, 2024). A vital misunderstanding of AI is, therefore, what makes it so dangerous; it should not be reduced to merely an online phenomenon created to aid with mundane human tasks

but instead be understood as a mechanism integrated into larger systems, from individual organisations to global supply chains, with real-world consequences (Stahl, Schroeder, Rodrigues, 2023).

This rise of technofeudalism has extreme implications for business and management. The demise of traditional capitalism, as argued by Varoufakis, has led to a drastic change in the way business is defined; the 'networked, top-down, mega-corporations we know today' have pushed 'the bakers, butchers and brewers of early capitalism to the sidelines' (Varoufakis, p.20). Extrapolating from Varoufakis's retailing example, a more profound inference can be drawn: Small and medium-sized businesses are now equally reliant on 'big techs' AI systems and, like individuals, have become serfs in a neo-feudalist system. In addition, ever-increasing reliance on the technology of these minority few companies has led to a phenomenon described as 'labour polarisation'. Technology and automation have led to the rise of a small minority of high-skilled jobs, a large majority of low-skilled jobs and a displacement of the jobs that were once in the middle (Zuboff, 2019). These issues directly compromise the UN's SDGs; decent work is becoming increasingly sparse, inequality is rising, and this shifting of power is breaking down traditional institutions. These failures can be directly attributed to AI's power; its ability to carry out surveillance, data extraction, and data interpretation for big tech firms has accelerated a shift away from the very ethics many of these companies were founded on. For example, Google's aims originally aligned with Kantian ethics; the algorithm learnt from users, and users learnt from the search engine, providing a closed feedback loop which was beneficial for all, rather than exploitative (Zuboff, 2019). However, as these companies realised the value of the data they were collecting, there was a collective shift; companies changed from mode 1 type

businesses, where profit was merely the means to social, moral and stakeholder success, to mode 2 type businesses, where profit became a sole objective (Kay, 1997). This collective shift has evolved how a vast majority of businesses are run; in this new world of extreme competition and profit chasing it has become increasingly challenging to maintain an ethical approach to business and management.

Arguably, the responsibility for changing the course of enterprises' toxic relationship with AI lies in the hands of business schools. Martin Parker (2018) challenges the assumption that business degrees should solely teach one form of organisation, capitalism, and instead suggests a radical reimagining of what business schools can be. If we are to 'reclaim humanity' from the grips of a technofeudal society run by a neoliberal elite and their ever more powerful AI systems, then business schools need to surpass the market-based, efficiency-focused teachings of Friedman and Taylor and adopt the socially responsible approaches of Rifkin and Raworth. Rifkin (2014) suggests that technological advancements have opened up opportunities for 'collaborative commons', whereby differing communities, businesses, and organisations can add value to one another whilst simultaneously adding value to themselves. Furthermore, he suggests it is not only what is taught in business schools that can and should change but how it is taught. The rise of technologies such as AI offers unprecedented opportunities for humanity; the collaboration enabled by these systems can change knowledge from its previous state as exclusive property to something of a publicly traded good, increasing global access to quality education.

Examples of radical management theories, such as Zohar's (2022) Quantum Management Theory, prove that a new age of thinking responsibly can work. QMT is based on moral principles, challenging traditional hierarchies of business by

championing ecosystems. These ecosystems, described by Zohar as complex adaptive systems (CADs), are 'holistic, constantly evolving and ... self-organising', and any imposition of hierarchies from the top 'alienates the constituent parts and destroys the creativity of the system' (ibid, p.41). When faced by the looming threat of AI, understanding businesses as natural systems is incredibly powerful; not only does it elicit a contrast between an ever-degrading natural ecosystem, often caused by the manufacturing and maintenance of these AI systems, but it also forces a consideration of the Kantian ethics of 'goodwill' that Vallor suggests are so important in the battle against AI's reflective nature. In practice, QMT has flourished; Haier's RenDanHeyi business model has gained international attention and acclaim. Haier's organisation into 4,000 ecosystem micro-enterprises and an offering of scenarios rather than products has not only led to commercial success, a revenue of \$28 billion in 2021 (Denning, 2022), but also social success; every one of China's 650,000 villages hosts a Haier community store, which also serves as a centre for training programs and after school care (Zohar, 2021). In the face of a looming 'technofeudal' crisis, business schools and businesses alike must adopt approaches such as QMT to reject the current neoliberal paradigm and move towards a more equal, inclusive and shared value future.

To conclude, and when considering the UN's SDGs specifically, AI has potential to both aid in immense good by providing education opportunities, health diagnoses and novel development ideas, or irreversible harm in the form of supporting a new age of 'cloud capitalism' and 'technofeudalism'. The path humanity takes will depend on how institutions react to threats posed by AI and whether they choose to adopt the morals and virtues suggested by Vallor or continue down a path of surveillance, data collection and privacy violations. Given that the exploitative systems of the past

are so ingrained into the AI tools of today, there is a vital necessity for a fundamental change in our interpretation of responsible business management; we must reject traditional hierarchies of neoliberalist capitalism and instead adopt radical new theories, such as Zohar's QMT.

Bibliography

Apple (2024), *Apple Intelligence Overview*, Available at:

https://www.apple.com/uk/apple-intelligence/?&mtid=20925rc739929&aosid=p238&mnid=swi0BE9IF-dc_mtid_20925rc739929_pcid_725095383037_pgrid_164170600955_pexid_ptid_kwd-342948744440_&cid=wwa-uk-kwgo-features--slid---productid--aibrand-aibrand-aibrand- (Accessed: 20/12/2024).

Brown, W. (2018) *In the Ruins of Neoliberalism*. New York: Colombia University Press.

Denning, S (2022) 'Can Firms Succeed Without Managers? The Case of Haier', *Forbes*, February 1st, Available at:

<https://www.forbes.com/sites/stevedenning/2022/01/30/can-firms-succeed-without-managers-the-case-of-haier/> (Accessed: 21/12/2024).

Elkington, J. (2018) 'The Triple Bottom Line: Here's Why It's Time to Rethink It', *Harvard Business Review (online)*, June 25. Available at:

<https://hbr.org/2018/06/25-years-ago-i-coined-the-phrase-triple-bottom-line-heres-why-im-giving-up-on-it> (Accessed: 19/12/2024).

Kant, I. (1997) 'Groundwork of the Metaphysics of Morals', in Gregor, M.

Timmermann, J. (ed.) *Texts in the History of Philosophy*. Cambridge: Cambridge University Press.

Kay, J. (1997) 'A Stakeholding Society: What Does it Mean for Business?', *Scottish Journal of Political Economy*, 44(4): 425-437.

Laney, D. (2023) 'AI Ethics Essentials: Lawsuit Over AI Denial of Healthcare', *Forbes*, November 16th. Available at <https://www.forbes.com/sites/douglaslaney/2023/11/16/ai-ethics-essentials-lawsuit-over-ai-denial-of-healthcare/> (Accessed: 17/12/2024).

Li, Y. & Shapiro, J. (2020) *China Goes Green*. Cambridge: Cambridge University Press.

McCallum, S. (2022) 'Meta settles Cambridge Analytica Scandal case for \$725m', *BBC*, November 23rd, Available at: <https://www.bbc.co.uk/news/technology-64075067> (Accessed: 20/12/2024).

McKinsey & Company (2024) *As AI advances from predictive to generative capabilities, its potential to address social issues defined by the UN Sustainable Development Goals expands*. Available at: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/ai-for-social-good> (Accessed: 19/12/2024).

Meta (2024), *We believe in the future of connection in the metaverse*. Available at: <https://about.meta.com/uk/metaverse/> (Accessed: 20/12/2024).

OpenAI (2024), *About*, Available at: <https://openai.com/about/> (Accessed: 20/12/2024).

Parker, M. (2018) *Shut Down the Business School: What's Wrong with Management Education*, London: Pluto Press.

Poole, E. (2024) *Robot souls: programming in humanity*, Boca Raton: CRC Press.

Raworth, K. (2018) *Doughnut Economics: Seven Ways to Think Like a 21st-Century Economist*, London: Penguin.

Rifkin, J. (2014) *The zero marginal cost society the internet of things, the collaborative commons, and the eclipse of capitalism*, New York: Palgrave Macmillan.

Stahl, B., Schroeder, D., Rodrigues, R. (2023) *Ethics of Artificial Intelligence: Case Studies and Options for Addressing Ethical Challenges*, New Jersey: Springer Cham.

Stiglitz, J. (2024) *The Road to Freedom: Economics and the Good Society*, London: Allen Lane.

Streek, W. (2016), *How Will Capitalism End? : Essays on a Failing System*, London: Verso.

United Nations (2024) *the 17 goals*. Available at: <https://sdgs.un.org/goals> (Accessed: 20/12/2024).

United Nations (2024) *AI for good*. Available at: <https://aiforgood.itu.int/> (Accessed: 19/12/2024).

UK Parliament (2024), 'What are the links between social media algorithms, generative AI and the spread of harmful content online?', Available at: <https://committees.parliament.uk/committee/135/science-innovation-and-technology-committee/news/203870/what-are-the-links-between-social-media-algorithms-generative-ai-and-the-spread-of-harmful-content-online/> (Accessed 20/12/2024).

Vallor, S. (2016) *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*, Oxford: Oxford University Press.

Vallor, S. (2024) *The AI Mirror: How to Reclaim Our Humanity in an Age of Machine Thinking*, Oxford: Oxford University Press.

Varoufakis, Y. (2023) *Technofeudalism: What Killed Capitalism*. Available at <https://www.vlebooks.com/Product/Index/3323318?page=0&startBookmarkId=-1> (Accessed: 20/12/2024).

Zohar, D. (2022) *Zero Distance: Management in the Quantum Age*, Singapore: Springer.

Zuboff, S. (2019) *The age of surveillance capitalism: the fight for the future at the new frontier of power*, London: Profile books.

Appendix

Appendix 1- Assignment Question

Part A. Explain Vallor's (2024) rationale for needing to "reclaim our humanity in an age of machine thinking", and how AI can be applied to the "restorative care and service of life" with examples; then consider the corporate challenges of deploying AI for good instead of reinforcing existing patterns of injustice, irresponsibility and unsustainability, along with possible alternatives should that fail, drawing on a range of relevant sources.

Part B. Critically evaluate the threat to humanity posed by "cloud capitalism" and "technofeudalism" (Varoufakis, 2023) and discuss the implications for business and management; then consider how business schools could help to avert this threat, e.g. by instigating a new paradigm for business and management no longer "dominated by last century's economic thinking and doing" (Raworth, 2018, p.292), drawing on relevant sources with examples.