Evaluating the Potential of AI for Good: Ethical AI and the Future of Sustainable Businesses

Allegra Hutchings

Durham University

Level 3, Business and Management

Category of Entry: Undergraduate students – text-based essays and reports

Word Count: 2989

Abstract

This essay utilises the United Nations Sustainable Development Goals (UN SDGs) to explore the ethical and corporate challenges of deploying artificial intelligence (AI) for good, in an era marked by rapid technological change and intensifying global crises. Building upon Shannon Vallor's (2024) call to 'reclaim our humanity', the integration of technomoral virtues in AI as a means of fostering collective well-being and sustainable development is evaluated. The tension between profit-driven motives and ethical standards further illustrates how unchecked usage of AI may reinforce systemic market and justice failures. This is seen in case studies such as the Post Office Horizon IT scandal and the exploitative practices of surveillance capitalism (Zuboff, 2019). The development of paradigms like cloud capitalism and technofeudalism (Varoufakis, 2023) concentrates economic power and wealth into the hands of a few and consequently undermines human autonomy and environmental sustainability. Emerging frameworks such as doughnut economics (Raworth, 2017) and quantum management theory (Zohar, 2022) could provide an alternative solution to entrenched neoliberal practices. Findings suggest that a cultural and institutional transformation is critical, with significant implications for policymakers and business schools, to reorient technological and economic systems towards a more sustainable and equitable future.

Contents

Introduction 4

Part A: Vallor's Rationale and Corporate Challenges 5

Vallor's Rationale for Reclaiming Humanity 5

Corporate Challenges in Deploying AI for Good 8

Part B: Threats of Cloud Capitalism and Technofeudalism 12

Cloud Capitalism, Technofeudalism and its Implications for Business and Management 12

Role of Business Schools in Averting the Threat 15

Alternative Paradigms 17

Conclusion 20

Bibliography 21

Appendix 24

Introduction

As technology advances and global crises intensify, humanity's use of artificial intelligence (AI) has been scrutinised. Shannon Vallor (2024) emphasises the need to 'reclaim our humanity' by integrating technomoral virtues into technology in order to benefit from AI's potential to contribute to the achievement of the United Nations Sustainable Development Goals (UN SDGs). However, systemic issues like cloud capitalism and technofeudalism (Varoufakis, 2023) threaten to perpetuate existing inequalities and biases. This analysis highlights the need for new approaches in business and management that prioritise long-term sustainability and ethical responsibility over short-term profit.

Part A: Vallor's Rationale and Corporate Challenges

Vallor's Rationale for Reclaiming Humanity

Shannon Vallor (2024) accentuates the necessity to 'reclaim our humanity' when integrating AI into society. The rationale for reclaiming humanity is centred on the notion that AI is used as a tool for beneficial collective growth rather than a mechanism for furthering unsustainable practices. Vallor argues that AI's development and application should be affiliated with technomoral virtues, underscoring the cultivation of ethical practices and human-oriented approaches when designing technology. However, as AI is "constructed entirely from the amalgamated data of humanity's past" (Vallor, 2024, p.6), its current trajectory risks ignoring human values by reinforcing biases in their data rather than fixing them, thus spreading inequalities and crises. Humanity is left with a choice: either continuously over-exploiting nature's resources to meet humanity's demands or ensuring our collective consumption is sustainable for current and future generations' well-being (Dasgupta, 2021).

AI's rapid development has triggered worries regarding existential crises – the most extreme being "the possibility of human extinction" (Vallor, 2024, p.162). These risks, however, extend beyond extinction – Vallor claims that an existential risk is "any threat to a future worth wanting" (2024, p.162). Vallor believes in 'technomoral wisdom' to address prevalent environmental and humanitarian concerns. Due to humanity's technomoral failures of current and past generations, it is now within our collective power to utilise AI ethically to stop these forces from consuming us.

If the triple bottom line is being met (ecological, social, and economic factors), AI could influence and contribute to the well-being of humanity and the planet, leading

to sustainable global development (Henriques & Richardson, 2004). Vallor advocates for AI to improve healthcare, advance sustainable resource management and address global inequalities, believing that some goals are “impossible to reach without advances in AI and related computing technologies” (2024, p.9). AI innovations aligned with agriculture, such as “precision agriculture and integrated pest management” (Dasgupta, 2021, p.491), could contribute profoundly to the reduction of our carbon footprint. AI is also integral to the improvement of workers’ productivity – a technological approach that complements employees as opposed to replacing them could help, for example, by collating more functional and informative information for human-decision making (Acemoglu & Johnson, 2023). Furthermore, the prospect of AI in healthcare is potentially ground-breaking, with machine learning already helping clinicians with diagnoses and rare genetic diseases (IEEE, 2019a). As “59% of UK parents” would already “seek out a doctor using AI for cancer diagnoses” (IEEE, 2019a), the prevalence of AI within healthcare and daily life will likely increase.

Despite clear evidence indicating technology’s transformative potential, as repeated claims suggest, benefits will only emerge if there are stark changes in how governments and corporations utilise AI alongside a “fundamental change in human consciousness” (Rifkin, 2014, p.333). Simply changing how AI is built and programmed is not enough, we need a “deeper and more lasting cultural transformation of our relationship to AI” (Vallor, 2024, p.11). Arguably, these are some of humanity’s gravest challenges, and the only way to tackle them is as a collective, with efforts “guided by practical wisdom” (Vallor, 2024, p.165). Making a change is not just for ourselves; we should constantly strive to “meet the needs of

the present without compromising the needs of future generations” (Roddick, 2000, p.56).

Corporate Challenges in Deploying AI for Good

Deploying ‘AI for Good’ faces significant corporate and structural challenges. This initiative, alongside ‘AI for SDGs’ created by China’s Academy of Science, aims at achieving the UN SDGs (Stahl et al., 2023). Vallor (2024) highlights a tension between profit-driven motivations and ethical imperatives where there is a prioritisation of shareholder returns over societal needs, noting that unchecked AI deployment risks reinforcing patterns of injustice, irresponsibility, and unsustainability. These initiatives have the potential to achieve SDGs by identifying disease patterns, predicting outbreaks, and improving the provision of public goods (IEEE, 2019b). Yet, their positive impact is limited by systemic barriers and profit-driven motives, such as efficiency over equity, highlighting the need for corporate responsibility and ethical oversight. For example, algorithms deployed to combat certain SDGs are exacerbating others, such as the high energy usage required for these technologies and the over-exploitation of resources (Stahl et al., 2023).

Although the market system is a principal driver of disparity, elucidated by Polman and Winston’s view that “climate change is the greatest market failure in history. Inequality is a close second” (2021, p.266), corporations must also focus on justice failure approaches. Businesses and corporations have an ethical obligation to use their influence to take responsibility for resolving justice failures (Singer, 2018). This underscores the notion that as AI technologies are mirrors of ourselves, the bias instilled in them, “whether it unjustly punishes us for our race, age, weight, gender…is not a computer problem. It’s a people problem” (Vallor, 2024, p.45) – drawing attention to surveillance capitalism’s exploitation of personal data and a

growing lack of autonomy (Zuboff, 2019). Surveillance capitalism monetises personal data without accountability, creating systems prioritising consumer manipulation over ethical use. This practice erodes trust, exacerbating systemic inequities. Privacy and data protection are human rights; therefore, AI's usage must be closely examined when reviewing ethical standards (Stahl et al., 2023). However, our society is focused on "material success" where "others are treated expediently and become reduced to instruments to accumulate more wealth" (Rifkin, 2014, p.313). Thus, to stop this reinforcement of biases, drastic change is imminent, both technologically and regarding humanity's patterns of consumption and production (Dasgupta, 2021).

The Post Office Horizon IT scandal is a poignant example of justice failure where algorithmic biases, flawed technology and institutional negligence led to wrongful convictions of sub-postmasters. This case, described as "the most widespread miscarriage of justice in English legal history" (Marshall, 2022, p.12) illustrates the dangers of technological opacity, lack of honesty and a corporate culture resistant to accountability. The Post Office's unwillingness to question the system's reliability and fear of reputational damage demonstrates how corporate irresponsibility intensifies systemic failures (Marshall, 2022). The scandal reflects neoliberal tendencies that prioritise efficiency and cost-cutting over ethics and transparency – arguably one of the root causes of corporate irresponsibility. The failure to address flaws in the Horizon IT system resulted in grave injustices, highlighting the importance of not only implementing technomoral wisdom in AI deployment (Vallor, 2024) but also the pertinence of new legislation and frameworks to address ethical management of technology and corporate responsibility.

We must move towards a sustainable mindset. Nevertheless, within many organisations and corporations, power differences are protected, causing decision-making to be susceptible to manipulation to preserve control (Stacey, 2003). Illuminated by Dasgupta (2021): “To shift consumption patterns in high-income countries, and the aspiring consumption patterns in the rest of the world, away from resource-intensive goods and services will require massive, coordinated actions” (p.236) – including governments and corporations of all sizes. As business is inherently unethical, it is imperative to embed technomoral virtues into AI development to ensure sustainable innovation (Vallor, 2024).

If these behaviours continue to persist with corporations failing to address sustainability and implement ethics voluntarily, Li and Shapiro (2020) suggest that the solution is ‘green authoritarianism’. Through top-down control, governments and institutions impose regulations to address ecological crises. China’s AI-driven initiatives for achieving the SDGs and The Green Silk Road Envoys programme align with green authoritarianism by addressing ecological challenges while fostering sustainable economic growth if it is “implemented with a rigorous focus on mitigating negative environmental side-effects and enhancing positive environmental services” (Li & Shapiro, 2020, p.123). Although green authoritarianism could deliver quick and significant ecological results, it raises ethical and political concerns by concentrating power in authoritarian institutions, creating opportunities for abuse; the corporate challenges for deploying AI for Good are intrinsically complicated. While AI offers powerful tools for fostering sustainability and justice, its deployment within existing

economic structures reinforces patterns of inequality and exploitation – challenges amplified by the rise of cloud capitalism and technofeudalism.

Part B: Threats of Cloud Capitalism and Technofeudalism

Cloud Capitalism, Technofeudalism and its Implications for Business and Management

The move towards cloud capitalism and technofeudalism threatens humanity's autonomy, equity, and sustainability. This exacerbates market and justice failures by concentrating power and wealth in the hands of a few, deepening social inequalities and environmental crises (IEEE, 2019b). Varoufakis (2023) identifies cloud capitalism and technofeudalism as emerging economic paradigms threatening humanity's social and economic foundations. For cloud capitalism, capital is embodied in data and algorithms and characterised by the dominance of digital platforms such as Amazon and Google. These platforms monopolise data and control the infrastructure on which businesses and individuals depend upon. Similarly, technofeudalism can be seen to amplify this power dynamic. With Varoufakis (2023) arguing that the "cost-of-living crisis that followed the recent pandemic cannot be understood outside the context of technofeudalism" (p.54), critics advocating for AI and cloud systems to be used for sustainable innovation must understand that these are currently perpetuating pre-existing instabilities, thus creating existential threats that Vallor forewarned about.

These systems where digital platforms govern economic activity place us in a "permanently active two-way street between our soul and the cloud-based system [allowing the] algorithmic network hiding in the cloud behind it, to guide our behaviours in ways superbly lucrative for its owner" (Varoufakis, 2023, p.31). These claims underscore the concept that individuals are sources of value extraction, with

every interaction supplying the cloud with personal data, enabling corporations to exploit this information for profit. While these cloud systems could be seen as more convenient and efficient, the loss of human agency and privacy from algorithmic decision-making threatens humanity's autonomy (IEEE, 2019b). Further aligning with Zuboff's (2019) notion of surveillance capitalism (a mechanism for control), these paradigms exacerbate market and justice failures by creating economic dependencies, undermining individual autonomy and societal welfare.

Companies like Google exemplify the dangers and potential of AI and data-driven business models. Although Google participates in initiatives like 'AI for SDGs', its core business model undermines these efforts by prioritising profit and data exploitation over systemic change (Varoufakis, 2023). Google's transition from a mode 1 to a mode 2 type company (Kay, 1997) has sparked concerns surrounding its covert data collection practices and monopolistic tendencies, raising ethical questions about user privacy, power imbalances and human behaviour exploitation for profit (Zuboff, 2019). As neoliberal policies aim to replace "regulation and redistribution with market freedom and uncompromised ownership rights" (Brown, 2019, p.58), deregulation and privatisation allow digital monopolies to thrive. Exemplified by 'tech giants' like Google, these algorithms and paradigms intensify what Raworth (2017) calls the 'treadmill of consumerism' – driving unsustainable production and consumption while failing to address global challenges.

As larger companies like Google dominate the market, smaller corporations are under pressure to adapt to AI's challenges. Businesses must embrace new management archetypes that prioritise technomoral virtues as globalisation means

“smaller local businesses can rarely compete with highly capitalised firms who seek market share” (Roddick, 2000, p.58). The long-term prediction that “smart IoT infrastructure…is going to carry on much of the economic activity” (Rifkin, 2014, p.301), accentuates the urgency for businesses and management to confront the ethical and societal implications of these technologies to evade AI becoming “the primary determinant” of future global economies (Vallor, 2024, p.176).

Role of Business Schools in Averting the Threat

Business schools can shape leaders and design systems prioritising human autonomy, data privacy, and ethical governance – challenging the exploitative tendencies of cloud capitalism and technofeudalism. To use 'AI for Good', institutional frameworks must be reconstructed, such as embedding ethical, sustainable, and inclusive theories into business curricula (IEEE, 2019b). Emphasising AI's potential, by "empowering the education sector with advanced courses on A/IS" (p.134) in undergraduate and postgraduate studies is essential when understanding AI's opportunities and risks in various contexts in the wake of the current global crisis (IEEE, 2019b). This is fundamental for the future of corporations and the wider public's perception.

A central issue in AI deployment is people being unaware of how their decisions are influenced by algorithms (Stahl et al., 2023). As "perception drives public response" (IEEE, 2019b, p.190), governments and organisations must have accurate knowledge of AI's potential to mitigate fear and increase awareness so that effective decision-making around emerging technologies transforms future society for the better (IEEE, 2019b). This understanding is most conducive when deployed from an early age. Dasgupta's (2021) findings that our connection with nature and the wider world "declines from childhood to an overall low in the mid-teens" (p.498) emphasise the urgency to instil education around ethics and sustainable management within schools.

Raworth (2017) advocates for doughnut economics – a framework critiquing cloud capitalism and technofeudalism's unsustainability by balancing human needs with

planetary boundaries. For Raworth, the concentrated focus on GDP growth has led to social inequality and environmental degradation – her circular-shaped model addresses the need to move beyond traditional economics. Incorporating quantum management principles further aligns with doughnut economics by fostering adaptability, collaboration, and ethical reasoning skills. If educational institutions implement these frameworks as well as embed courses on ethics and technomoral virtues to the same extent as their current investment in profitable gains, there is hope that current and future generations could navigate modernity's complexities (Vallor, 2024).

The Post Office scandal highlights the importance of ethics and accountability training. Business schools could use this case study as a cautionary example, emphasising the importance of responsible technology deployment, management, and development. Incorporating frameworks like doughnut economics and quantum management theory alongside case studies will accentuate the necessity of fostering ethical innovation aimed at achieving the UN SDGs. Utilising both cautionary (Horizon IT scandal) and exemplary case studies (Haier) in education will show why these approaches to leadership, focussing on collective well-being rather than individual profit, are crucial.

Alternative Paradigms

The debate on rethinking economic theories and structures has shifted towards more radical, multidimensional concepts, focussing on transforming the neoliberal growth paradigm. As neoliberalism “has taken us to the brink of ecological, social and financial collapse” (Raworth, 2017, p.12), stronger regulation and oversight is necessary, with governments enforcing stricter antitrust laws to prevent monopolistic behaviour by tech giants. As Raworth and Dasgupta suggest, transitioning economies and corporations away from GDP growth to collective well-being is essential for long-term sustainability.

Stakeholder capitalism is an ideology that can further long-term sustainable business practices by focusing on value creation for all stakeholders rather than just shareholders (Philips, Freeman & Wicks, 2003). Stakeholders lack power within exploitative businesses, raising questions of who speaks for marginalised groups (the environment and those of a lower socio-economic background). The notion of voice is paramount and could address issues regarding transparency (Stahl et al), therefore programming humanity and Vallor’s technomoral virtues into AI is critical.

Quantum management theory aligns with stakeholder capitalism’s ideology, providing a practical framework for rethinking organisational leadership and decision-making (Zohar, 2022). Human interaction and effective communication within corporations are essential for employees and customers (IEEE, 2019b), and quantum management emphasises this holistic thinking, interconnectedness, and technomoral wisdom. This new managerial paradigm contrasts traditional top-down corporate structures and neoliberal business practices that prioritise short-term

profits over long-term societal benefits. Governments and organisations could adopt quantum-inspired regulatory models that are adaptive, responsive and aligned with the UN SDGs.

The Haier Group is a Chinese multinational known for its innovative ‘Rendanheyi’ management model. ‘Rendanheyi’ emphasises decentralisation, employee empowerment and customer-focused operations (Zohar, 2022). Haier’s practices illustrate forward-thinking management approaches, underpinning ethical innovation, adaptability, and stakeholder-focused business. Haier’s emphasis on empowering employees and fostering creativity supports quantum management theory, and thus Vallor’s appeal for technomoral virtues. Haier’s decentralised structure enables inclusive decision-making, mitigating the risks of reinforcing existing biases in AI development. This business model provides a counterexample to the hierarchical structures of cloud capitalism and technofeudalism. Its approach fosters agility and equitable value distribution, showcasing how businesses can operate ethically and sustainably within a competitive global economy (Zohar, 2022).

Transitioning to new paradigms could be challenging due to entrenched interests and philosophies. The West is described as an “individualistic society”, using a more “analytical and hermeneutical approach” (IEEE, 2019b, pp.203-204), with a focus on the application of frameworks and theories rather than the embedding and teaching of morals. Whilst Eastern philosophical ethics remain influential by offering an alternative to Western individualistic frameworks, the IEEE (2019b) states the importance of returning to “normative foundations” (p.205) to address reliable innovation. In bridging Western and Eastern philosophies by focusing on shared

themes of virtue cultivation, relational ethics, and adaptability, Vallor's technomoral wisdom, together with 'RenDanHeyi' offers a framework that can be applied to various emerging theories to tackle concerns around AI and sustainability, requiring both individualism and collective harmony (Zohar, 2022).

Conclusion

AI's rapid development, alongside the rise of cloud capitalism and technofeudalism, represents a profound threat to humanity, worsening inequality, ecological degradation, and systemic exploitation. Vallor's (2024) technomoral virtues stress AI's transformative potential when aligned with justice and sustainability. However, as demonstrated by Google and the Post Office, systemic market and justice failures remain significant barriers to deploying AI for Good. Business schools have the potential to counter this by cultivating new generations equipped with ethical and sustainable frameworks such as doughnut economics and quantum management theory that incorporate technomoral virtues. Any methods or tools used to achieve one SDG should not exacerbate another, therefore systemic change requires more than educational reform – robust regulatory interventions and collectively reimagining economic paradigms within existing corporations are needed (Vallor, 2024). By implementing modifications within companies and education, humanity can avert existential threats posed by these emerging economic systems and benefit from AI's transformative potential throughout society.

Bibliography

Acemoglu, D. & Johnson, S. (2023) *Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity.* Great Britain: Basic Books.

Brown, W. (2019) *In The Ruins of Neoliberalism: The Rise of Antidemocratic Politics in the West*. New York: Columbia University Press.

Dasgupta, P. (2021) *The Economics of Biodiversity: The Dasgupta Review*. Available at: https://assets.publishing.service.gov.uk/media/602e92b2e90e07660f807b47/The_Economics_of_Biodiversity_The_Dasgupta_Review_Full_Report.pdf [Accessed: 20 December 2024].

Henriques, A. & Richardson, J. (eds) (2004), *The Triple Bottom Line: Does It All Add Up.* Oxford: Taylor & Francis Group. Available from: ProQuest Ebook Central [Accessed: 16 January 2024].

IEEE (2019a) *IEEE Unveils Third Annual 'Generation AI' Global Study, Illuminating Trust Millennial Parents Have in Artificial Intelligence for the Health and Wellness of Their Generation Alpha Kids.* Available at: https://www.ieee.org/about/news/2019/ieee-unveils-third-annual-generation-ai.html [Accessed: 21 December 2024].

IEEE (2019b) *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*. Version 2. Available at: https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf [Accessed: 21 December 2024].

Kay, J.A. (1997) 'A Stakeholding Society: What Does It Mean for Business?', *Scottish Journal of Political Economy*, 44(4), pp.425-435.

Li, Y. & Shapiro, J. (2020) *China Goes Green – Coercive Environmentalism for a Troubled Planet*. Cambridge: Polity.

Marshal, P. (2022) 'Scandal at the Post Office: The Intersection of Law, Ethics and Politics', *Digital Evidence & Electronic Signature Law Review*, 19, pp.12-28.

Philips, R., Freeman, E. & Wicks, A. (2003) 'What Stakeholder Theory Is Not', *Business Ethics Quarterly*, 13(4), pp.479-502.

Polman, P. & Winston, A. (2021) *Net Positive: How Courageous Companies Thrive by Giving More Than They Take*. Boston, Massachusetts: Harvard Business Review Press.

Raworth, K. (2017) *Doughnut Economics: Seven Ways to Think Like a 21st-Century Economist.* London: Random House Business Books.

Rifkin, J. (2014) *The Zero Marginal Cost Society: The Internet of Things, the Collaborative Commons, and the Eclipse of Capitalism.* New York: St. Martin's Press. Available from: ProQuest Ebook Central [Accessed: 20 December 2024].

Roddick, A. (2000) 'Beautiful business', *RSA Journal*, 148(5495), pp.54-59.

Singer, A.A. (2018) *The Form of the Firm: A Normative Political Theory of the Corporation.* New York: Oxford Academic. Available from: https://doi.org/10.1093/oso/9780190698348.003.0013 [Accessed: 21 December 2024].

Stacey, R. (2003) *Strategic Management & Organisation Dynamics* (4th Ed). Harlow: FT Prentice-Hall.

Stahl, B.C., et al. (2023) *Ethics of Artificial Intelligence: Case Studies and Options for Addressing Ethical Challenges.* Cham: Springer International Publishing.

Vallor, S. (2024) *The AI Mirror: How to Reclaim Our Humanity in an Age of Machine Thinking.* New York: Oxford Academic. Available from: https://doi.org/10.1093/oso/9780197759066.003.0008 [Accessed: 21 December 2024].

Vallor, S. (2016) *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting.* New York: Oxford Academic. Available from: https://doi.org/10.1093/acprof:oso/9780190498511.003.0011 [Accessed: 21 December 2024].

Varoufakis, Y. (2023) *Technofeudalism: What Killed Capitalism.* London: Vintage Digital.

Zohar, D. (2022) *Zero Distance: Management in the Quantum Age*. Bern: Springer Nature. Available from: ProQuest Ebook Central [Accessed 21 December 2024].

Zuboff, S. (2019) *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*. London: Profile Books.

Appendix

Original assignment Brief:

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.

Answer both parts. Each part carries equal weight.

Overall word limit: 3000 words