Responsible AI: Developing Artificial Intelligence with Moral Courage

AP Seminar

14th April 2024

Word Count: 2186

Innovation and Moral Responsibility in Artificial Intelligence

Just two months after its launch, ChatGPT, the popular artificial intelligence software, reached 100 million monthly active users, making it the fastest-growing consumer application in history (Hu, 2023). Furthermore, Bloomberg predicts the generative AI market to grow by over 3200% to a market size of $1.3 trillion over just the next decade (2023). This unparalleled growth of artificial intelligence has the potential to revolutionize every aspect of our lives, from enhancing human efficiency to pushing the boundaries of technology and innovation. However, the rapid rise of artificial intelligence also comes with many potential consequences. For instance, the automation of jobs will displace millions of workers, particularly widening the gap between the rich and the poor (Georgieva, 2024; Bankins & Formosa, 2023). Similarly, the energy consumption and carbon footprint associated with large AI models also pose significant sustainability concerns (Chien et al., 2023; Li et al., 2023; Wu et al., 2021). Margaret Goralski, a professor at Quinnipiac University, describes AI as a double-edged sword, stating that although it holds immense promise for increasing sustainability and growth, it also poses significant social, economic, and environmental challenges in the global economic order of today (Goralski & Tan, 2020). Essentially, with AI being developed in a capitalist society, where short-term gains often overshadow ethical considerations, organizational culture and operational pressure can cause the values of developers to blur, especially when the mission becomes more important than virtues (Thomas & Chaleff, 2017). With the widespread influence of artificial intelligence, this short-term “ends justify the means” mindset could be catastrophic. This begs the question: how can AI leaders ensure

their pursuit of innovation does not overshadow their moral responsibility to mitigate its negative impacts on society? A potential solution to this dilemma is adopting a perspective of moral courage in the development of artificial intelligence. As defined by Dr. Ted Thomas, the Director of the U.S Army CGSC, moral courage describes the willingness to stand up for what is right, even in the face of societal disapproval and personal cost (2017). An example of this moral courage can be seen in the actions of Simone Biles, who chose to prioritize her mental health over competition, showcasing a different kind of bravery where she stood up for something more important than the continuous pursuit of excellence (Thompson II, 2021). Similar to how Simone Biles showed courage by stepping back from competition to prioritize her mental health, AI leaders must exhibit moral courage by regulating AI development to ensure ethical integrity, even if it might slow down innovation or reduce short-term profits. This prioritization of morality over innovative success is essential for AI leaders to navigate the ethical dilemmas present in the growth of artificial intelligence. To achieve this, AI leaders must institute legislation to ensure ethical transparency, establish retraining programs to maintain economic inclusivity, and make advancements in green technology to promote environmental sustainability.

Transparency and Trust in AI Development

A significant concern with the growth of artificial intelligence is the lack of transparency in its development process, which introduces risks like embedded biases and systemic discrimination. Most of the leading organizations in the development of

artificial intelligence do not have open-source models, transparent documentation, or any information related to the collection, training, and development of their AI models. A team of researchers from MIT, Stanford, and Princeton developed a transparency index for the top 10 AI developers — including OpenAI, Google, and Meta — that evaluated over 100 indicators like data source selection and personal information usage. Based on this index, the highest-scoring developer only scored points for 54 of the 100 indicators, with the average score for all developers only being 37 (Bommasani et al., 2023). This 'blackbox' nature of AI makes it impossible to audit the large amounts of data that these models learn from, making it difficult to trust AIs and eliminate embedded biases, while also raising serious concerns about privacy and consent (Bankins & Formosa, 2023). In fact, in an attempt to conceal the racial biases embedded in Google's Gemini model, the company ended up developing a model that inadvertently amplified racial discrimination by refusing to generate images of people of certain races (Robertson, 2024). Evidently, the "risks associated with AI have already begun to compound on top of existing inequalities, resulting in further harm to already marginalized groups" (*UNESCO Ethics of Artificial Intelligence*, 2022). With AI growing to be such a widespread technology, AI leaders must understand that tackling the dangerous potentialities of artificial intelligence is more important than continuing to push for innovation and adoption. Although such transparency is a frightening prospect in a capitalist economy, it is essential for AI leaders to prioritize the welfare of society over individual gain and glory (Thompson II, 2021). Leading AI organizations must shift towards greater transparency and actively engage in creating mechanisms for accountability and ethical governance. Specifically, AI leaders

must “develop auditing mechanisms…to identify unwanted consequences, such as unfair bias”, and “develop appropriate legal procedures [to] improve the digital infrastructure of the justice system” (Floridi, 2023; Khogali & Mekid, 2023). Ultimately, by fostering a culture of transparent and ethical AI development, organizations can ensure that their innovations continue to push the boundaries of innovation while also upholding essential ethical considerations.

Economic Equity with the Growth of AI

Similar to the ethical concerns caused by the growth of artificial intelligence, the global economic landscape will also face significant disruptions. Although artificial intelligence serves to enhance human efficiency and unlock new opportunities, it can also disrupt the existing job market and worsen the existing economic divide. Kristalina Georgieva, the managing director of the International Monetary Fund, predicts that “in advanced economies, over 60 percent of jobs may be impacted by AI… which could lower labour demand, leading to lower wages and reduced hiring” of lower-skilled workers (2024). Similarly, Dr. Sarah Bankins, a professor with expertise on the economic effects of artificial intelligence, agrees with Georgieva, reiterating that “AI use may disproportionately and negatively affect lower-skilled and lower-paid workers, while its benefits may disproportionately accrue to those with higher skills and wages” (Bankins & Formosa, 2023). Not only does this imply increasing automation in the job market, but it also suggests a shift in the very nature of human work, employment stability, and income levels. Furthermore, this disruption is not uniform and will tend to negatively affect

workers who are already in lower-income brackets. Beyond the individual level, the IMF also assessed the readiness of over 125 countries, finding that wealthier economies tend to be better equipped for AI adoption than low-income countries (Georgieva, 2024). While artificial intelligence will help rich nations and high-income individuals to leapfrog ahead, it will also expand the economic gap and leave millions of low-income individuals and underdeveloped nations struggling to catch up. To combat these challenges, AI leaders must work towards making artificial intelligence more accessible and equitable. Georgieva states that "it is crucial for countries to establish comprehensive social safety nets and offer retraining programs for vulnerable workers" (2024). Essentially, by investing in retraining programs and educational initiatives, AI leaders can ensure that the displaced workforce can adapt to the changing job market and develop skills that enable them to work better using AI. Not only does this make the AI transition more inclusive, but it also ensures that the economic benefits of AI are more evenly distributed amongst society. Similarly, Dr. Luciano Floridi, the Founding Director of the Digital Ethics Center at Yale University, suggests that AI leaders must "develop legal instruments... to lay the foundation for a smooth and rewarding human-machine collaboration in the work environment" (2023). By fostering a culture of inclusivity and economic equity, AI leaders can ensure that the growth of artificial intelligence protects jobs and does not worsen the divide between the wealthy and the poor.

Sustainable Development Practices

Accompanying the ethical and economic challenges of artificial intelligence is the significant environmental impact of AI technologies. From electricity for powering computers to water required for cooling data centers, the resources consumed by AI models are already staggering. Now, with the rapid growth of artificial intelligence and higher predicted growth of adoption, its negative effects on the environment are of concern. According to a paper authored by lead researchers at Meta, Facebook reported a 2.9× and 2.5× capacity increase for AI power usage over the recent 18 months, highlighting the staggering environmental impact of AI (Wu et al., 2021). Similarly, Microsoft's water usage spiked by over 30% to 1.7 billion gallons of water in 2022 (*2022 Environmental Sustainability Report*, 2022), which can be attributed to the tremendous growth of their partner, OpenAI. Likewise, Google's water usage also increased by 20% in 2022, withdrawing 25 billion litres… of potable water… for cooling data centers" (Li et al. 2023). Clearly, the environmental impact of artificial intelligence is a concern that must be addressed to ensure the sustainability of AI technology in the future. Therefore, AI leaders must work towards developing capable AI models without escalating environmental costs. Dr. Andrew A. Chien, a professor of computer science at the University of Chicago, suggests the adoption of the CarbonMin algorithm, which reduces carbon emissions by 35% by directing artificial intelligence queries to data centers in low-carbon regions (Chien et al., 2023). Similarly, Meta’s Carole-Jean Wu noticed reductions in emissions by optimizing models, infrastructure, and hardware (2021). Ultimately, it is evident that AI leaders must "take a deliberate approach when developing

AI technologies [by] considering the environmental impact of innovations and taking a responsible approach to technology development" (Wu et al., 2021). By adopting such sustainable practices, AI leaders can ensure that the growth of artificial intelligence is environmentally sustainable and does not worsen the existing climate crisis.

However, as another potential solution, many researchers suggest that artificial intelligence can optimize other existing technologies to reduce emissions and promote sustainable development around the world. For instance, Dr. Margaret Goralski states that "AI-enabled applications... have already enhanced the efficiency of industries [and] helped to conserve precious, non-renewable resources [that contribute] to global sustainability". Similarly, other researchers believe that AI will have a significant social impact on sustainable development, climate change, and environmental concerns... [resulting in] cleaner, less polluted, and more liveable cities" (Khogali & Mekid, 2023). Essentially, these researchers believe that AI has the potential to revolutionize the way we approach sustainability, making it a powerful tool for addressing environmental challenges. Regardless, while AI may enhance efficiency and resource conservation, the immediate environmental cost is unsustainable. In fact, research states that "the global AI demand may be accountable for 4.2 – 6.6 billion cubic meters of water withdrawal in 2027, which is more than the total annual water withdrawal of 4 – 6 Denmarks" (Li et al., 2023). Furthermore, advancements in AI models have not yet reached a stage where they can offset their consumption significantly. For that reason, Carole-Jean Wu, the director of AI research at Meta, reiterates that "advancing the field of machine intelligence must not in turn make climate change worse", and that AI must be developed with a deeper

understanding of its environmental implications (Wu et al., 2021). Essentially, while AI may have the potential to optimize existing technologies and promote sustainable development, it is essential to recognize that these solutions are merely temporary band-aids for the problem, and the real solution lies in developing AI technologies that are sustainable from the outset.

Leading Innovation with Ethics and Sustainability

The growth of artificial intelligence presents many potential ethical, economic, and environmental challenges that must be addressed to ensure a sustainable future. Ethically, the lack of transparency in AI development poses significant risks, from embedded biases to overall distrust. Similarly, the job market and the economy may also face disruptions, with AI automation potentially displacing millions of workers and widening the economic divide. Furthermore, the environmental impact of AI is already staggering, with the carbon and water footprints already posing significant sustainability concerns. Ultimately, AI leaders must not only push the boundaries of innovation but also ensure their developments are built on a foundation of ethicality and sustainability. By fostering a culture of transparent AI development, organizations can uphold essential ethical considerations while still making strides in technological innovation. By investing in retraining programs and educational initiatives, AI leaders can ensure that artificial intelligence is equitable and accessible to all, protecting jobs and preventing the widening of the economic divide. Additionally, by prioritizing sustainability and funding research into more environmentally friendly AI technology, leaders can mitigate the negative

environmental impact of AI growth. Essentially, the growth of artificial intelligence must be guided by moral courage — a commitment to standing up for what is right, even in the face of societal disapproval and deferred innovation. As stated by Ted Thomas, "[moral courage] is the one essential, vital quality of those who seek to change the world", and it often "just takes one person to take a stand and bring the voice of reason and light into a dark room" (2017). For the future of AI to be viable, AI leaders must step forward with moral courage, ensuring that their innovations are not only groundbreaking but also sustainable and ethical.

Works Cited

2022 Environmental Sustainability Report. Microsoft. (2023, May 10). https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW15mgm

Bankins, S., & Formosa, P. (2023). The ethical implications of Artificial Intelligence (AI) for Meaningful work. *Journal of Business Ethics*, *185*(4), 725–740. https://doi.org/10.1007/s10551-023-05339-7

Bloomberg. (2023, June 1). *Generative AI to become a $1.3 trillion market by 2032*. Bloomberg.com. https://www.bloomberg.com/company/press/generative-ai-to-become-a-1-3-trillion-market-by-2032-research-finds/

Bommasani, R., Klyman, K., Longpre, S., Kapoor, S., Maslej, N., Xiong, B., Zhang, D., & Liang, P. (2023). The Foundation Model Transparency Index. *ArXiv*. https://doi.org/10.48550/arXiv.2310.12941

Chien, A. A., Lin, L., Nguyen, H., Rao, V., Sharma, T., & Wijayawardana, R. (2023). Reducing the carbon impact of generative AI inference (today and in 2035). *Proceedings of the 2nd Workshop on Sustainable Computer Systems*. https://doi.org/10.1145/3604930.3605705

Floridi, L. (2023). *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford University Press.

Georgieva, K. (2024, January 14). *AI will transform the global economy. let's make sure it benefits humanity.* International Monetary Fund. https://www.imf.org/en/Blogs/Articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity

Goralski, M. A., & Tan, T. K. (2020). Artificial Intelligence and Sustainable Development. *The International Journal of Management Education*, *18*(1). https://doi.org/10.1016/j.ijme.2019.100330

Hu, K. (2023, February 2). *ChatGPT sets record for fastest-growing user base*. Reuters.com. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

Khogali, H. O., & Mekid, S. (2023). The blended future of automation and AI: Examining some long-term societal and ethical impact features. *Technology in Society*, *73*. https://doi.org/10.1016/j.techsoc.2023.102232

Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models. *ArXiv*. https://doi.org/10.48550/arXiv.2304.03271

Robertson, A. (2024, February 21). Google apologizes for 'missing the mark' after Gemini generated racially diverse Nazis. *The Verge*. Retrieved April 10, 2024, from https://www.theverge.com/2024/2/21/24079371/google-ai-gemini-generative-inaccurate-historical.

Thomas, T., & Chaleff, I. (2024). Moral Courage and Intelligent Disobedience. *AP Seminar Performance Task 2: Individual Research-Based Essay and Presentation - Directions and Stimulus Materials*, 8–16. The College Board.

Thompson II, M. (2024). Simone Biles and the Most Human Meaning of Courage. *AP Seminar Performance Task 2: Individual Research-Based Essay and Presentation - Directions and Stimulus Materials*, 41-44. The College Board.

United Nations Educational, Scientific and Cultural Organization. (2022). *Ethics of Artificial Intelligence*. UNESCO.org. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

Wu, C., Raghavendra, R., Gupta, U., Acun, B., Ardalani, N., Maeng, K., Chang, G., Behram, F.A., Huang, J., Bai, C., Gschwind, M.K., Gupta, A., Ott, M., Melnikov, A., Candido, S., Brooks, D., Chauhan, G., Lee, B., Lee, H.S., Akyildiz, B., Balandat, M., Spisak, J., Jain, R.K., Rabbat, M.G., & Hazelwood, K.M. (2021). Sustainable AI: Environmental Implications, Challenges and Opportunities. *ArXiv*, https://doi.org/10.48550/arXiv.2111.00364

Works Referenced

Brynjolfsson, E., Li, D., & Raymond, L. (2023). Generative AI at Work. *National Bureau of Economic Research*. https://doi.org/10.3386/w31161

Clarke, R. A., Schot, J., Fischer, K., Gray, R., Wells, R. P., Piet, J., Smart, B., Esty, D. C., Cairncross, F., Bavaria, J. L., Greeno, J. L., & Stavins, R. N. (1994, July). *The Challenge of Going Green.* Harvard Business Review. https://hbr.org/1994/07/the-challenge-of-going-green

Hao, K. (2020, December 7). *Training a single AI model can emit as much carbon as five cars in their lifetimes*. MIT Technology Review. https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for Deep Learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. https://doi.org/10.18653/v1/p19-1355

Sustainable AI: Environmental Implications, Challenges and Opportunities. (2022). *ArXiv*. https://doi.org/10.48550/arXiv.2111.00364