**Safety First Responsible AI: An Ethicist’s Standpoint**

Confucius’ Custodians

**Introduction**

Ethics are principles that guide human behavior; they are foundational principles to approaching issues that affect life and global stability. They are ideals such as equity, safety, and fairness that create a framework for how AI systems should work in a broad sense and provide end goals to achieve. Ethical principles are at odds with effective accelerationism due to rapid AI development and lack of oversight, regulation, and stakeholder input. Without critical analysis of these systems, we risk destabilizing economies, environmental damage, and societal values such as privacy. AI must be responsible and created in an ethical manner with so much on the line.

**Reckless AI Development Undermines Progress, and Leads to Substantial Macro-Level Impact**A safety-first approach to AI development is not just wise but is imperative for ensuring stable, equitable, and sustainable progress during this transformative era. The breakneck speed of AI advancement presents significant risks to global employment, economic stability, and societal cohesion. The International Monetary Fund warns that up to 40% of global jobs could be impacted by AI, with advanced economies facing particularly acute disruptions in sectors like finance, healthcare, and administration (Kristalina, 2024). Without immediate interventions such as large-scale reskilling programs and robust social safety nets, we risk plunging millions into long-term unemployment and exacerbating economic polarization.  
 Accelerationists argue that rapid AI deployment drives innovation, but history teaches us that reckless speed often results in catastrophic consequences. Look no further than the Three Mile Island nuclear incident: a failure to prioritize safety in a high-stakes technology led to widespread fear, economic loss, and a necessary overhaul of industry practices (World Nuclear Association, 2022). Do we really want to repeat such mistakes with a technology as pervasive as AI, one that directly shapes the global economy, governance, and human lives?  
 Unchecked accelerationism also concentrates wealth and power, deepening inequality. As per Goldman Sachs, AI could boost global GDP by 7% annually over a decade, but without equitable distribution mechanisms, these gains will benefit only a select few—large corporations and developed nations—while sidelining smaller businesses and less developed economies (Goldman Sachs, 2023). This is not innovation; it’s exploitation masked as progress. Further, accelerationism’s neglect of ethical safeguards risks embedding systemic biases at a global scale. Poorly regulated AI systems in hiring, lending, and other critical domains have already amplified discrimination. Imagine the societal harm if such biases are scaled up through rapid, unregulated deployment. Thus, as ethicists, we must reject this reckless gamble and demand AI development that serves humanity, not just profits or political power.

**Rapid Development of AI Has a Detrimental Effect on the Environment**Training large-scale AI models requires immense amounts of computational power, resulting in significant energy consumption. Since the early 2010s, the computational power needed to train state-of-the-art AI models has been doubling approximately every 3.4 months (Kanungo, 2023). Some studies suggest that generating a single image with AI can consume energy equivalent to charging a smartphone or 1000x that of what would be required to generate text outputs (Crownhart, 2024). AI systems rely heavily on extensive data center infrastructure for processing and storage.   
 The construction of these data centers to support the AI models involves materials that are severely carbon-intensive, such as steel, cement, and semiconductor chips. According to the UN Environment Programme (2024), these centers rely on extensive electronic infrastructure that demands vast amounts of raw materials. For example, producing a 2 kg computer requires approximately 800kg of raw materials, illustrating the significant resource consumption involved.  
 Additionally, manufacturing AI hardware, such as GPUs and TPUs, is both energy and water-intensive. Semiconductor production is a critical step in creating these components and requires enormous quantities of ultrapure water for cleaning silicon wafers. The mining and refining of rare earth metals used in AI hardware further strain global water resources, adding to the environmental burden.  
 Data centers, which are essential for AI operations, have emerged as the leading cause of water shortages in the state. Leading to a disruption in the needs of the local community. In fact, Virginia’s water use in this region rose by nearly two-thirds, from 1.13 billion gallons in 2019 to 1.85 billion gallons in 2023 (Hodgson, 2024). The growing water demands and environmental challenges posed by AI technologies underscore the urgent need for the tech industry to adopt greener, more sustainable practices.

**Accelerated AI Will Weaken Society and Humanistic Capacities**With AI systems sophisticated to dystopian extents, data is collected about individuals, their preferences, behaviours and emotions (Mario Fialho, 2023). This massive surveillance and collection of data raises serious questions about the right to privacy for our team.  
 AI generates decisions based on processes run over data. Essentially, it predicts next steps based on history. The biggest advances in our race have come when we learnt from history, and worked that learning into today’s context. We build on historic knowledge - AI may be doing that better. But we also build on socio-cultural context, and intangible concepts like “instinct”, “leap in the dark” and “failure sets”. We attempt crazy, unimagined and unpredicted things - only because we are not AI systems. From recommendation algorithms that shape our viewing habits to AI-driven political campaigns that sway voter opinions, these systems are increasingly making decisions that were once the exclusive domain of human beings (*5 AI Ethics Concerns the Experts Are Debating | Ivan Allen College of Liberal Arts*, n.d.). As we become more reliant on AI-powered tools and platforms, we risk surrendering our agency and critical thinking skills to algorithms that may not always have our best interests at heart. This general shift under accelerated AI not only threatens privacy and the barrier between digital and real but also risks atrophying important cognitive abilities and problem-solving skills that have been central to human progress throughout history.   
 With natural language processing increasingly advancing, it doesn’t seem like there will be much left for human-to-human connection in our lifestyles. You couldn’t possibly joke and snicker with the robot who serves you your drive-through order about yesterday’s game like you would with someone real serving it in the same jersey as you. Accelerated AI threatens to diminish the richness and nuance of human relationships.  
 Fairness and accountability is no easy play for either humans or AI systems. For humans though, there is ‘context’ to a judgement, and there is ‘instinct’ about a business investment – that AI systems won’t have. These inscrutable black-boxes have been seen as perpetuating bias and discrimination (The Princeton Review, 2024). The opacity of these systems makes it challenging to identify and rectify errors, resulting in life-altering consequences for individuals subjected to AI-driven decisions - usually also marginalized.

**Challenges and Trade-offs**  
Ethics do not offer set guidelines for how to build systems or products that will do no harm or function to reach specific end-goals of reducing social inequalities. Instead, they offer a “spirit” in which our actions should align with. As such, ethicists face several challenges and must make trade-offs in multiple scenarios such as in competitive markets, on the societal and geopolitical scales, and within the legal system as more and more AI systems are built that create social, environmental, and personal challenges.  
 First, human ideals that are hard-coded into ethics such as fairness, truth, and trust are difficult to achieve because there is no holistic definition that can be applied cross-culturally and cross-contextually (Larsen & Dignum, 2024). Second, organizations often prioritize rapid deployment over safety to capture market share. Third, regulatory bodies lag behind technical advancements and policymakers often lacking domain expertise.  
 Ethicists must concede that they do not have the answers to all of these distinct, nuanced problems but instead offer end goals that should be pursued when trying to come to equitable solutions with other stakeholder groups.

**References**

*5 AI Ethics Concerns the Experts Are Debating | Ivan Allen College of Liberal Arts*. (n.d.). Iac.gatech.edu. https://iac.gatech.edu/featured-news/2023/08/ai-ethics

Crownhart, C. (2024, May 23). *AI is an energy hog. This is what it means for climate change.* MIT Technology Review. https://www.technologyreview.com/2024/05/23/1092777/ai-is-an-energy-hog-this-is-what-it-means-for-climate-change/

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

Goldman Sachs. (2023). *Generative AI could raise global GDP by 7%*. Goldmansachs.com. https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent

Hatzius, J., Briggs, J., & Pierdomenico, G. (2023). *The Potentially Large Effects of Artificial Intelligence on Economic Growth (Briggs/Kodnani)*. Gspublishing.com. https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45d7-967b-d7be35fabd16.html

Hodgson, C. (2024, August 18). *Financial Times*. @FinancialTimes. https://www.ft.com/content/1d468bd2-6712-4cdd-ac71-21e0ace2d048

Kanungo, A. (2023, July 18). *The Green Dilemma: Can AI Fulfil Its Potential without Harming the Environment?* Earth.org. https://earth.org/the-green-dilemma-can-ai-fulfil-its-potential-without-harming-the-environment/

Larsen, B., & Dignum, V. (2024, October 17). *AI value alignment: Aligning AI with human values*. World Economic Forum. https://www.weforum.org/stories/2024/10/ai-value-alignment-how-we-can-align-artificial-intelligence-with-human-values/

Mario Fialho. (2023, June 19). *The Moral and Ethical Implications of Artificial Intelligence*. Stefanini. https://stefanini.com/en/insights/articles/the-moral-and-ethical-implications-of-artificial-intelligence

The Princeton Review. (2024). *Ethical and Social Implications of AI Use | The Princeton Review*. Www.princetonreview.com. https://www.princetonreview.com/ai-education/ethical-and-social-implications-of-ai-use

UN Environment Programme. (2024, September 21). *AI Has an Environmental problem. Here’s What the World Can Do about that.* UNEP. https://www.unep.org/news-and-stories/story/ai-has-environmental-problem-heres-what-world-can-do-about

World Nuclear Association. (2022, October 11). *Three Mile Island Accident - World Nuclear Association*. World-Nuclear.org. https://world-nuclear.org/information-library/safety-and-security/safety-of-plants/three-mile-island-accident

**Appendix**

Interactions with LLMs-  
  
Input prompt:   
  
*Polish the tone for the excerpt below:   
Training large scale AI models requires substantial computational power, leading to a high energy consumption, to levels unimaginable (MIT Technology Review). Some sources show that the energy it takes to generate an image is equivalent to that of charging a phone. AI systems require extensive data center infrastructure for processing and storage. These facilities generate substantial heat during operation, necessitating efficient cooling systems.; not to mention the infrastructure that would be needed to support these AI models necessitates building data centers, which rely on materials that are severely carbon-intensive, such as steel, cement, and, of course, semiconductor chips. Many data centers rely on water-based cooling, consuming large volumes of water daily. This demand intensifies water scarcity in drought-prone regions, creating competition between industrial use and local communities. Virginia has suffered water shortages, especially in the areas where these data centers are clustered, such as Northern Virginia. This is not coincidental, as data centers-a backbone for AI operations-are big users of water for cooling. In fact, water use by data centers in Virginia's "Data Center Alley" grew almost two-thirds from 1.13 billion gallons in 2019 to 1.85 billion gallons in 2023. (Financial Times) Moreover, manufacturing AI hardware, including both GPUs and TPUs, is an extremely energy and water-intensive process. As an example, semiconductor manufacturing-a key process in creating these components-requires immense volumes of ultrapure water for cleaning silicon wafers. Also, the mining and processing of rare earth metals used in AI hardware further contribute to global water demand and, therefore, to the environmental impact. The impact of AI on the environment is not restricted to water usage. E-waste generated through AI systems is usually dumped improperly, releasing toxic elements into the water bodies, hence polluting water. In addition, such water released by cooling systems at a very high temperature causes thermal pollution, which bothers the aquatic ecosystem and disturbs biodiversity. The increasing water needs of AI and related environmental impacts encourage the technology sector to go green.*  
  
Output:  
*Here’s a polished version with a refined tone and improved flow:*

*Training large-scale AI models requires an immense amount of computational power, resulting in significant energy consumption—levels previously unimaginable (MIT Technology Review). Some studies suggest that generating a single image with AI can consume energy equivalent to charging a smartphone.*

*AI systems rely heavily on extensive data center infrastructure for processing and storage. These facilities produce substantial heat during operation, necessitating the use of efficient cooling systems. Moreover, the construction of data centers to support these AI models involves materials that are highly carbon-intensive, such as steel, cement, and semiconductor chips.*

*Many data centers utilize water-based cooling systems, consuming vast amounts of water daily. This demand exacerbates water scarcity in drought-prone regions, creating competition between industrial requirements and local community needs. For instance, in Virginia—particularly in Northern Virginia’s "Data Center Alley"—water shortages have been an increasing concern. Data centers, which are vital to AI operations, are significant water consumers. In fact, their water use in this region rose by nearly two-thirds, from 1.13 billion gallons in 2019 to 1.85 billion gallons in 2023 (Financial Times).*

*Additionally, manufacturing AI hardware, such as GPUs and TPUs, is both energy- and water-intensive. Semiconductor production, a critical step in creating these components, requires enormous quantities of ultrapure water for cleaning silicon wafers. The mining and refining of rare earth metals used in AI hardware further strain global water resources, adding to the environmental burden.*

*The environmental impact of AI extends beyond water usage. Improper disposal of e-waste generated by AI systems releases toxic elements into water bodies, polluting ecosystems and harming biodiversity. Moreover, high-temperature water discharged from cooling systems causes thermal pollution, disrupting aquatic ecosystems and further disturbing biodiversity.*

*The growing water demands and environmental challenges posed by AI technologies underscore the urgent need for the tech industry to adopt greener, more sustainable practices.*

Input:  
*This demand exacerbates water scarcity in drought-prone regions, competition between industrial requirements and local community needs. Virginia is considered water rich.*   
  
*Emphasize and highlight the fact that datacentres are the leading cause for the water shortages.*

Output:   
*This demand places an enormous strain on water resources, even in regions traditionally considered water-rich, such as Virginia. Data centers, which are essential for AI operations, have emerged as the leading cause of water shortages in the state. For example, in Northern Virginia’s "Data Center Alley," water usage by data centers grew nearly two-thirds, from 1.13 billion gallons in 2019 to 1.85 billion gallons in 2023 (Financial Times). This significant consumption is directly tied to their reliance on water-based cooling systems, intensifying competition between industrial needs and local communities, and pushing the region closer to water scarcity despite its historical abundance.*

Reflection:   
GPT-4o helped greatly in setting the tone(one that is similar to my tone, but without severe pain staking amounts of time and effort) of the argument and for highlighting certain points. Helped ensuring grammatical correctness while being succinct.