The World I Want

The real debate about artificial intelligence is not whether it will shape the future but whether people will have the ability to influence the direction it takes. The history of the internet and smartphones reminds us how quickly technology can change society, creating opportunities while also enabling hacking, surveillance, and widespread misinformation. AI is moving along this same trajectory but at a much faster pace. For that reason, I believe it is essential to imagine the kind of world I want to live in and to define the ethical commitments and governance structures that can help bring it into being. My independent study in AI ethics and governance is meant to strengthen both my technical expertise and my capacity to contribute to frameworks that ensure AI development respects human values.

The world I envision is one where people remain central to decision-making and where machines serve to extend, rather than diminish, human dignity and fairness. Already, many critical judgments in hiring, finance, education, and healthcare are filtered through algorithmic systems that often remain opaque. A job applicant may never know why their résumé was rejected, or a patient might be unaware that an algorithm determined their eligibility for treatment. Such cases reveal how the absence of transparency undermines trust and accountability. In the society I want, individuals should always have the right to understand how important decisions were reached, to question them, and to know how their personal information is used. These protections are not luxuries, they are necessary rights in a world where automated systems play increasingly decisive roles (Eubanks, 2018).

Fairness must also be more than a rhetorical commitment. Algorithmic bias has repeatedly been shown to disproportionately harm already marginalized groups, particularly Black and brown communities. Cathy O'Neil's description of "Weapons of Math Destruction" captures how opaque, large-scale models can produce systemic inequality when left unchecked (O'Neil, 2016). Ethical AI design requires fairness to be integrated at every

stage of system development from how data is collected to how models are evaluated. It also requires that AI tools undergo cultural adaptation so that they do not impose majority values on diverse communities. Without that, technology risks becoming a new form of cultural imperialism. For me, fairness means not only auditing error rates but also asking whose experiences are left invisible and who is excluded from conversations about design.

Safety and accountability are equally critical. It is not enough for companies to make voluntary promises about responsible AI. History has shown that when profit motives conflict with ethical responsibility, corporations often prioritize speed and market dominance. Real accountability must come from enforceable regulation. International policy experiments are already underway under the European Union's AI Act, the U.S. Executive Order on Safe, Secure, and Trustworthy AI, and the NIST AI Risk Management Framework all represent important first steps (*Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, 2023) (White House, n.d.) (Council of the European Union, 2024). These frameworks introduce obligations for risk assessments, audits, documentation, and bans on harmful practices. The world I want takes these experiments further, creating binding treaties, independent ethics councils, and global enforcement mechanisms similar to those governing nuclear technology.

The definitional debate over what AI "really is" illustrates why adaptive governance is so important. While some have argued that systems like large neural networks genuinely exhibit intelligence and others say sophisticated pattern matchers without understanding. Whether AI understands is an interesting philosophical question, but governance cannot wait for consensus. Instead, regulations should focus on observable properties such as scale, autonomy, and social impact. A famous metaphor of AI governance as driving a car captures this well, we do not need perfect foresight to drive safely, but we do need a clear windshield (transparency), good steering (adaptive policy), and reliable brakes (safety mechanisms). These principles will remain relevant regardless of how AI capabilities evolve.

My personal reasons for pursuing this study are deeply connected to my background. Growing up in Zimbabwe, I experienced firsthand how technological inequality can widen opportunity gaps. I also recognize that the communities least represented in AI policymaking are often those most impacted by algorithmic decisions. As a researcher, I am drawn to areas like bias in natural language processing, compliance, and automated reasoning because they provide opportunities to build systems that are both technically sound and socially accountable. I see technical accuracy and ethical responsibility as inseparable stronger documentation and interpretability not only make models easier to maintain but also create pathways for oversight and redress.

The goals of my independent study align with this vision. I want to deepen my technical literacy in how AI models are designed and deployed, while also exploring ethical frameworks that help clarify competing values. Beyond that, I aim to engage with real-world policy tools, such as audits and compliance processes, to understand how they function in practice. Case studies ranging from AI in hiring and education to surveillance and national security will serve as concrete examples to analyze risks and governance gaps. I also intend to create applied outputs, such as an ethical impact assessment for a specific AI application, complete with harm analysis, stakeholder mapping, and governance recommendations.

The readings that guide my study underscore why urgency is needed. Works like Automating Inequality and Algorithms of Oppression expose how data-driven systems perpetuate injustice, while The Age of Surveillance Capitalism critiques how corporate incentives push relentless extraction of personal data (Noble, 2018) (Zuboff, 2019)(Eubanks, 2018). There is a lot of concern against scaling without accountability .Contemporary policy texts such as the EU AI Act and the NIST AI RMF outline possible regulatory pathways. Collectively, these sources remind me that AI governance is not an abstract academic exercise, it is an urgent societal necessity.

Ultimately, the world I want is not defined by fear or naïve optimism. It is a world where AI amplifies human creativity, strengthens institutions, and

distributes benefits fairly, rather than reinforcing hierarchies or eroding rights. Achieving this world requires sustained study, deliberate design, and courageous governance. As a student and future researcher, my task is to contribute to this effort by combining technical insight with ethical responsibility. If we succeed, AI will not control us, it will instead reflect a future we intentionally shaped together.

References

Artificial Intelligence Risk Management Framework (AI RMF 1.0). (2023, January 1). NIST Technical Series Publications. Retrieved August 31, 2025, from https://doi.org/10.6028/NIST.AI.100-1

Council of the European Union. (2024, 07 12). *European Union. (2024). Regulation laying down harmonised rules on Artificial Intelligence (AI Act). Official Journal of the European Union*. European Union. http://data.europa.eu/eli/reg/2024/1689/oj

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor.* St Martins Press.

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown.

White House. (n.d.). *Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence.*

https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.