Dasha Asienga
Mathematics and Statistics Department – Statistics Honors Student

Unveiling Bias: Towards Fairness in Data-Driven Decision-Making

What if your college admissions decision was determined by an artificial intelligence (AI) model? What if your prospective employer relied on AI to make your hiring decision? In an increasingly data-driven world where decision-making is automated, these scenarios aren't just hypothetical – they're slowly becoming a reality that we face. From healthcare to criminal justice to facial recognition to advertisement to credit and banking, these models wield immense power. But lurking behind the scenes of these algorithms lies the scary, hidden truth: many of these models are biased and unregulated, and they reinforce and even propagate societal discriminatory practices. This can lead to exacerbating disparities, especially among marginalized groups.

My thesis delves into this pressing issue, seeking to define unfairness in a mathematically and statistically tractable way and explore ways to mitigate the bias embedded within such predictive models, with a particular focus on the Seldonian framework. Without diving into too much technicality, Seldonian algorithms allow us to produce models that guarantee fairness, as defined, with high confidence (probability). However, this is not achievable without trade-offs. As such, my thesis aims to study the far-reaching implications of employing the Seldonian framework. Acknowledging that theory by itself is not enough, my thesis culminates with a practical application of these insights to a real-world data set used in risk assessment tools throughout the United States to determine bail and parole decisions. The implications of my research are profound, paving the way for more equitable use of machine learning and AI models in almost every domain. By mitigating algorithmic unfairness, we can foster equity in crucial decision-making processes.

My thesis fits the 3MT model because it focuses on a complex topic with significant real-world implications. Such algorithms are pervasive in every aspect of life and affect us all: these models score teachers and students, sort resumes, grant (or deny) loans, evaluate employees, target voters, determine parole, and monitor our health. I look forward to being able to effectively engage with the audience and communicate the relevance, urgency, and importance of my research in just three minutes.