Dasha Asienga

3MT Thesis Competition Speech

AI, artificial intelligence, do you trust it? Honestly, I get a little scared. And I'm not talking about ChatGPT or self-driving cars. I'm talking about the kind of AI that replaces human judgment to decide whether you'll land that consulting job after graduation, whether your sibling will get into their dreaM college, and whether you'll get a promotion or lose your job.

Currently, in the criminal justice system, an algorithm called COMPAS is used to predict whether defendants will reoffend in the future. This tool helps determine whether they should be granted bail or parole(!). It seems like AI would make decision-making so much faster. But, there's an issue.

COMPAS is twice as likely to label as "high-risk"(/), Black defendants, who we know did not reoffend, compared to White defendants who also didn't reoffend. It also makes the opposite mistake; it's more likely to label White defendants who reoffended as low risk, compared to Black defendants who also reoffended!

COMPAS! It's supposed to guide us in the right direction(!), yet it exhibits algorithmic bias. This happens when a model systematically discriminates against a demographic group, even when we don't feed it demographic information! This can perpetuate, and amplify, existing human biases. That's where *my* statistics thesis comes in.

First, I needed to define what fairness means mathematically. For me, what was most important is making sure my model is equally wrong or equally right for all groups. I don't want a model to over predict that a male applicant, who is just as qualified as me, is more suitable for a job.

I then used a novel technique known as the Seldonian frameworK, to constrain my model to satisfy that fairness definition, within some margin, probability, and confidence level. I fit this framework on the COMPAS data set and ran simulation studies to test this idea to its limit, and I found that, with some substantial trade-offs, we *can* have fairer AI models!

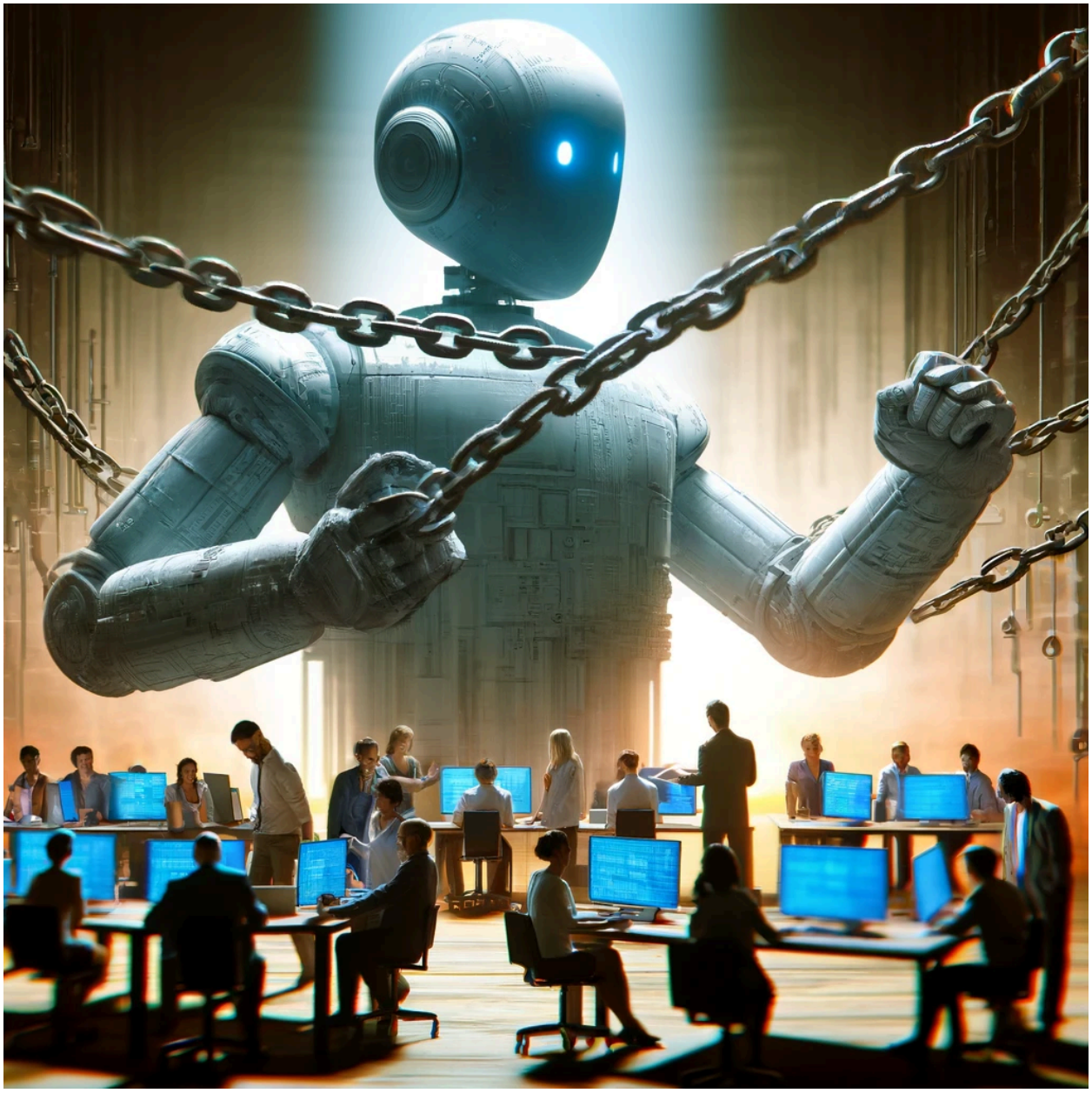
But the solution to fair AI is not going to be solved by computer scientists, data scientists, and statisticians alone. Everyone in this room has a role to play. We need domain experts in every field to define what a fair AI landscape looks like for their application area. But we also need to look inwardly and address the biases we have as humans in society, trying our best not to pass that bias on to the data in the first place.

My research is a part of ongoing work that gives us hope for the future where a lot of our decision-making will be automated. It gives us hope for a more just world!

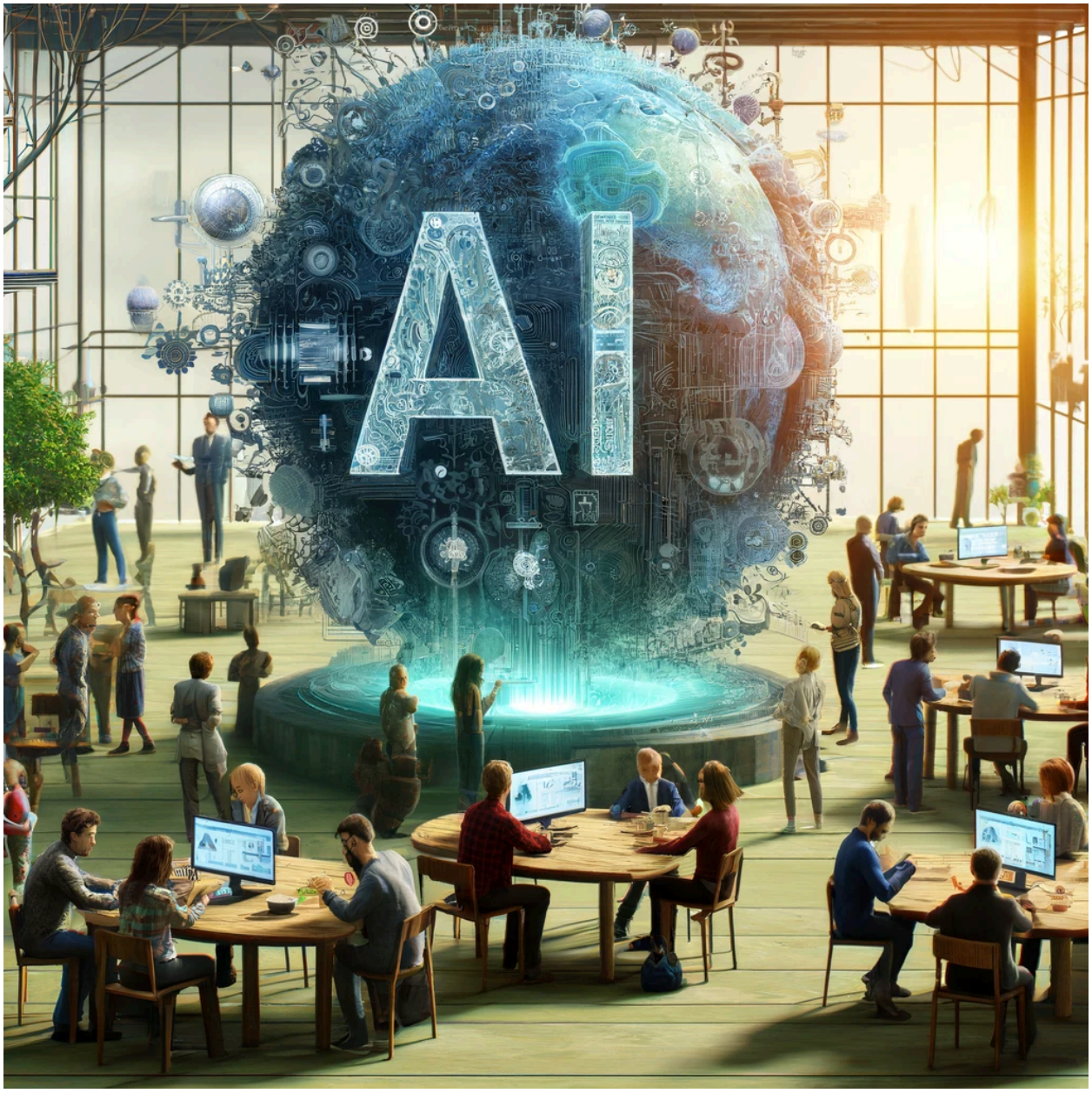
* underlined key words are to be emphasized during speech.

Title Ideas (Short & Memorable):

1. Unveiling Bias: Toward Fairness in Data-Driven Decision-Making
2. Decoding Bias: Toward Fair AI
3. **Decoding Bias: Can AI be Fair?**
4. Decoding Algorithmic Bias: A Path to Fair AI



(generated by AI, no pun intended haha)



(generated by AI, no pun intended haha)