

Areas of Risk	Steps To Address Bias, Discrimination, and Unfairness
Non-Representative Training Data	<p>Within any project a company interacts with AI, the question of “What are we doing” must be asked. Along every step of the way that the AI you are training is being developed this question and the notion of “Do No Harm” must be at the top of mind. I believe that in order to release an AI model out to the public some level of understanding about the model, what it is trained on, and why it generates the results that it does, must be understood. It is for this reason that managers and employees must all act in a manner that the product they are building is not simply to generate revenues for their companies, but instead that it meant to be a tool that lifts up humanity and most certainly does no harm. The data that our AI models are fed need to represent a diverse background of different cultures, beliefs, perspectives, and languages. Simply feeding the model only one narrative leaves it to be susceptible to unintended bias, miss representation, and an uncomplete picture to the actual reality of certain situations. When addressing this issue within the work place, it needs to be understood that this particular area is fundamental within the training of AI, and that if it is clear the models are beginning to not represent reality or have any regard for underrepresented populations, the model immediately needs correction.</p>
AI and Violence	<p>Another fundamental area within the creation and monitoring of AI models is an absolute zero tolerance policy for models that project violence. In absolutely no situation is okay for a model to actively encourage someone to take preform acts of violence against anyone else. If a released model is caught doing this in any capacity it needs to be immediately suspended and addressed. This is issue also can not be one that is ignored in less represented areas, languages, and communities. No matter the location of the user, the language being used, or the prompt being inputted, the return from a AI model should never be encouraging violent acts to be performed. Teams need to work to monitor the activity of AI across all areas of its use, not just the most mainstream areas, in order to ensure that this level of absolute zero tolerance for this type of language is never generated.</p>
Sourcing and Sustainability	<p>One of my majors here at DeSales is Supply Chain Management, where we speak a lot about this particular issue. When addressing this issue with likes of AI being involved within the process as well, much more needs to be considered. AI that is capable of searching market prices for potential vendors and business partners needs to be trained on the core competencies of the business as well. Things such as never dealing with companies that practice child labor, poor worker safety conditions, or any illegal activity what so ever needs to be incorporated into the model. The model must return suggestions for sourcing that is not only concerned with profitability but also at a bare minimum meeting these base requirements as well.</p>
Fairness in Loan Programs	<p>My other major here at DeSales is Finance. Through some of the courses I have taken within this major, I have heard and began to understand the usage of AI and computers being used to screen candidates and determine their credit risk. This is a very similar thing as AI’s reading over job applicants resumes and denying some before an actual human even interacts with the application. If I ever find myself within a position where I am apart of a team in charge of, or interacting with, one of these machines, there are a few key things I will have to made sure of. First is</p>

	<p>that something such as race or ethnicity should never be something that we allow our models to make these decisions based off of. Other things like a person's gender should also be only taken into consideration when absolutely necessary to make a fair prediction towards someone credit risk. Limiting these things and having models focus on other variables is key in order to provide fair and effect loans to our customers. We must at all stages treat our model with great respect and concern knowing that it has great power and can change the lives of peoples and businesses based off of the outputs it gives us. With this in mind using AI in this setting needs to be taken very seriously with lots testing and monitoring in place. A human element should also be present, and we should always respect a customer's desire to have a human evaluate their request for a loan if they so wish for that.</p>
Insider Trading	<p>Another major problem that could be present with AI is if it ever is permitted to make stock market trading decisions for companies and their investments. If this is the case then the AI is going to need to be well defined into what some illegal activities are and that it may never come in contact with. Things like insider trading could protentional be something an AI pick up information on and attempt to then make business decision based off of. Any information that could be deemed insider information will need to be taken extremely seriously and not be fed into the AI modal under any circumstances to ensure that this issue never becomes a reality for a company.</p>