AI for High-Stakes Decision-Making

Hima Lakkaraju

Decision makers such as judges make crucial choices affecting human lives on a daily basis. Such *high-stakes* decisions have a significant and lasting impact on individuals as well as society. However, prior research has demonstrated that the judgments made by these decision makers are not only inconsistent and error-prone but also subject to a variety of biases - e.g., racial and gender biases. Can we help these decision makers make better judgments?

Over the past few years, I have been focusing my research efforts on finding answers to the above question. My fascination with this question began during the early years of my Ph.D. at Stanford when I got a chance to discuss about the pitfalls of human decision-making in criminal justice with my advisors Prof. Jure Leskovec, Prof. Sendhil Mullainathan, Prof. Jon Kleinberg, and Prof. Jens Ludwig. This discussion inspired me to explore how Artificial Intelligence (AI) can help human decision makers. We quickly realized that there are certain fundamental challenges that hinder the applicability of existing AI techniques to improve high-stakes decision-making: a) The available data only captures the outcomes of the decisions made by human decision makers and not the counterfactuals. b) The data is prone to selection biases and confounding effects. c) The successful adoption of AI in high-stakes decision-making relies heavily on how well decision makers can understand and trust its functionality; however, most of the existing AI models are are not very interpretable.

I collaborated with an interdisciplinary team of economists, statisticians, and computer science researchers from Stanford, Cornell, Harvard, Duke, and Microsoft Research to develop novel AI frameworks which address the aforementioned challenges, there by, paving the way for large-scale deployment of AI to aid high-stakes decision-making. I proposed novel formulations and optimization procedures to learn AI models which are both accurate as well as interpretable. The proposed approaches were designed to account for missing counterfactuals and selection biases in the data. At the core of these approaches lie rigorous techniques spanning various topics in AI literature such as Bayesian modeling, Markov decision theory, multi-objective optimization, submodular optimization, and multi-armed bandits.

Policy simulations on a dataset of bail decisions of 758K defendants in New York City demonstrated that our AI frameworks reduce the crime rate by up to 24.8% with no change in jailing rates, or alternatively reduce the jailing rates by 42.0% with no increase in crime rates. Similar results were observed on data from 40 other large urban counties in the US. Our methodology is also effective in aiding other high-stakes decisions such as treatment recommendations in healthcare, hiring, and strategic decisions in business. The aforementioned findings received widespread media attention and were discussed in various outlets including MIT Technology Review, The New York Times, and Harvard Business Review.

Our work has had significant academic as well as real-world impact. The above research has been published in premier AI conferences (NIPS, AISTATS, KDD, AAAI, SDM), top journals in economics (Quarterly Journal of Economics) and psychology (Psychological Methods), and has received multiple best paper awards including the INFORMS best data mining paper prize (finalist). I was also named one of the 35 innovators under 35 by MIT Tech Review. Our frameworks are currently being used by Montgomery school district for aiding school officials in identifying at-risk students. Our algorithms, results, and insights are also being leveraged by certain court systems and insurance companies.