Yale University

Prof. Nisheeth K. Vishnoi
Department of Computer Science   *Campus Address:*
P.O. Box 208285   *Arthur K. Watson Hall*
New Haven, Connecticut 06520-8285 *51 Prospect St.*
New Haven, CT 06511

Email: *nisheeth.vishnoi@yale.edu*
Website: *http://www.cs.yale.edu/homes/vishnoi/Home.html*

Date: September 30, 2019

Subject: *Support for S.1876/H.2701, A commission to ensure good government in the age of artificial intelligence*

Dear Senator Pacheco, Representative Gregoire, and members of the Committee:

I am writing this to offer strong support for S.1876 and H.2701, *An Act Establishing A Commission On Transparency And Use Of Artificial Intelligence In Government Decision-Making* and *An Act Establishing A Commission On Automated Decision-Making, Artificial Intelligence, Transparency, Fairness, And Individual Rights*. I am a Professor of Computer Science at Yale. My research focusses on designing algorithms. For the last few years, I have studied ways in which problems can arise when algorithmic or machine learning (ML) solutions are applied on humanistic or societal contexts and, in parallel, worked on finding solutions to these. Many of my works have led to publicly available tools that have the ability to control discrimination in a variety of situations. Demos of some of these are available from my website[1].

It is becoming crucial that governments, policymakers, and the public to understand the range and extent of the problems that can be caused due to the deployment of algorithms in society. Today's society consists of humans living in a complex and interconnected world that is intertwined with a variety of computing, sensing and communicating devices -- generating massive amounts of data that is systematically stored. Artificial intelligence (AI) systems, powered by algorithms that are capable of learning from this data are driving how humans interact with each other (e.g., Facebook, Twitter), interact with information (e.g., Google), conduct business (e.g., financial trading, sharing economy platforms such as Uber) and learn (e.g., YouTube or other educational technology). On the one hand, the ever-increasing speed of computing devices, connectivity of the Internet, the Internet of things, and the sophistication of algorithms comes with the promise of immense prosperity. On the other hand, these advances allowed algorithms to look deep into human behavior and decision-making processes and, as a consequence, such automated entities have gained the ability to nudge human behavior and possibly even govern it. Several recent events suggest that algorithms can be discriminatory, reinforce human prejudices, polarize opinions, accelerate the spread of misinformation, and are generally not as objective as they are widely thought to be.

---

[1] http://cs.yale.edu/bias/

Many of the aforementioned problems arise because the algorithms driving the corresponding technology have been designed to maximize utility functions and one or more links in the cyclic interactions between humans, data, models, algorithms, and humans is broken -- whether it is how and what human-generated data is being captured and presented to algorithms, or the model choices not being aware of issues such as discrimination and privacy, or the impact algorithmic outputs can have on human behavior. Thus, requiring transparency about the algorithm's objectives and the underlying data sets is crucial in increase the trust of the people, and also limiting the power of private corporations who can use these to further their goals.

Data collection methodologies and frameworks (from surveys, police reports, to curriculum vitaes) largely assume that the information in them will be processed by intelligent humans that have the capacity to deliberate. On the other hand, in a world where humans interact with technology, data about humans is being recorded in formats designed by computer scientists keeping in mind issues such as efficiency, and a lot is lost in translation. As an example, consider the case of data being recorded by police officers in the stop-and-frisk programs in NYC. This data is now being used by machine learning algorithms, that might not have all the context a human would have, to learn and give rise to automated tools with the goal to do the same job. While the intention is to increase transparency and reduce human-bias in such decision making, there is nothing to prevent these machine learning algorithms to learn from and amplify human biases.

Discrimination, fairness, trust, etc. represent some of the most basic human values, they are incredibly complex, nuanced, cultural, and context dependent. In many cases, social/legal scholars have to argue for lengthy periods of time to come to a conclusion if a particular act was fair in a particular context or not. While scientists are working towards broadening computational models to be more versatile and flexible, and designing algorithms with the awareness that their outcomes may affect humans and society at large, current AI/ML systems are very far from incorporating this breadth and nuance of human knowledge. incapable of adapting to a specific context.

Overall, AI/ML tools can negatively affect people and society in potentially discriminatory and irreversible ways and should be extremely carefully evaluated before government or public use. Thus, I strongly support this legislation that recommends the formation of a commission that surveys the government's use of algorithmic decision systems, makes recommendations to ensure that the adoption of new technologies does not harm individuals or socially salient groups, and inform the general public about the current uses of AL/ML systems in Massachusetts state government operations.

I remain at your disposal if you need more information.

Nisheeth K. Vishnoi