Dr Michael G. Kollo

Qurious Analytics

Dear Sir / Madam,

Firstly, let me praise your efforts here to create a thoughtful document and tackle some very difficult issues around society, ethics, technology and artificial intelligence. I have a deep interest in this area, both on a personal and professional front, and it is something that I have been writing and speaking about for some years. I am primarily a technical leader, so my primary strength is understanding exactly how these algorithms create forecasts of various characteristics, and how they intersect (or not) with human decision making under uncertainty. I have a strong academic background (PhD from London School of Economics), and a long career in senior management in financial services in London and globally. Many of my comments below will be more from a technical feasibility point of view, as well as some practical applications.

I have contacted your office directly in the past to collaborate if you required this kind of input, and I extend that invitation again.

Please find my responses below, and I hope to be in touch later.

Question	Comments
A	I think here the 'similarly significant' is the weakest part of this definition. Any kind of 'profiling' that is undertaken, if that is for offer of services/goods, or for scholarships, or jobs, or a constraint like non-approval of a Visa or insurance for example, are mostly commercial targeting outcomes. They may not be legal, but they can constrain or minoritise groups. Increasingly difficult is where this effect is incremental across multiple algorithms, where each algorithm carries a small bias, but collective they create enough of one to be a concern. As a note: as a critical starting-point, do you define 'significant' decisions as those that are identified at a micro (individual) level or those that create macro (society) level impacts. Human rights are defined at the micro level, but ultimately speak to the shape of society they create. Many cases these two can conflict.
B	The rebuttable presumption should also consider what the explanation would have been without that AI information, as many of these decisions made today with limited quantitative input, or basic analytics, do not explain how these decisions came about.
C	Generally, yes. Transparency that does not infringe intellectual property is desirable, and ideally, there should be publicly available white papers for all algorithms that 'touch' more than a certain number in the population.
D	No, I don't think so. I think dutiful care should be taken in decision making (with or without AI), but the law should not push individuals to indiscriminately intervene or override. In my experience, the

	reason to introduce AI into decision making is not that it automates existing human intuition, but that it is superior to it.
E	I think that certain algorithms are going to relevant for human rights, but many won't be. For proposal 14 to have 'bite' ideally, the first problem to solve is to identify the correct algorithms that impact human rights, and follow this with the kinds of considerations regarding bias and fairness primarily. The toolkit then helps developers that work in that area, eg profiling students for scholarships, profiling citizens for criminal activity, to closely consider the complexities of their specific problems when it comes to human rights. Ideally, many of these decisions already exist in our society, but are qualitatively undertaken. I would suggest that the guidelines that frame the qualitative decision making should be the ones that also frame the algorithmic models. In that sense, it is not the creation of new information, but rather 'bring to the front' the existing ethical considerations that may be latent or assumed.
F	The regulatory sandbox is an excellent idea. Technically, it should be a series of tester algorithms (ethical adversarial networks) that test the output of an algorithm for bias, fairness and report these. One has to be careful as these are just indications, and should not be held up in court as legal basis for fairness. The other way, is not to build the algorithms, but actually run competitions/commercial propositions for private enterprise to create systems for evaluation of various fairness criteria and make these available in a single framework.
G	Startups could be very effective in this space, and perhaps having an accelerator programme sponsored and housed by the government would ensure that these companies receive the funding and help they need to get to a decent scale.
H	The online courses relating to data science and coding practises should incorporate something around human Rights, but again, would require more specialisation in terms of the kinds of algorithms, and where they are appropriate.

I would add two primarily comments on your excellent paper, and I am happy to discuss further if required.

1. The problem of scale. Algorithms and AI are often seen as dangerous for governments, regulators and society because they make forecasts and actions at enormous scale. While we can be happy that an individual may make biased/unfair decisions, they are only a single person, and in a democracy and capitalism, we believe that many decision makers converge to a mean or average judgement that represents our values. In the case of algorithms, it is equivalent to a dictatorial decision maker: a single system for evaluation, and this makes us nervous. We should absolutely focus our attention on those algorithms that are mass-scaled firstly, so anything that touches the lives of 20,000 or more people as a rough example. From an individual level, human rights are agnostic to scale, but at a society (top-down) level algorithms that impact more of us in a similarly unfair way, for example, are much more dangerous than smaller pockets. There are technical solutions to this (ensemble modelling for example) where no single model makes the decision. Something to consider.

2. The taxonomy of decisions. I think there is a required piece for your proposal which creates much more depth in terms of the types of decisions that are assisted by algorithms (so offer of service, eg sales, or a denial of a service, or a profiling for monitoring etc), which has information about what kind of information is used, and importantly for what purpose. I think that would greatly help the practical application of much of what you have referred to, and ensure that it is targeted specifically into pockets of society where it matters. Not having this framework I think leaves you at a risk of being very broad in your statements, and missing the bridge from 5,000 feet (principles kind of statements) down to the 500 feet applications (ok, but exactly which tests should I care about).

I have many more thoughts on the subject, but I suspect that you will receive many different comments from many different agencies, each on what is seen as vital for them. In my case, I have no vested commercial interest in promoting any of the above comments, but I have a deep personal believe that this is an area where we collectively need to 'get right' to ensure a smooth transition to a wide-scale responsible adoption of AI.

--

Michael G. Kollo, PhD

CEO and Founder

Qurious Analytics

w: www.quriousanalytics.com

Qurious Analytics Pty Ltd |