CCC
Computing Community Consortiun
Catalyst

AAAI/CCC Symposium on AI for Social Good

Overarching Discussion

Amulya Yadav: We're going to start the final thing on our agenda, except for the plenary talk, which is the Overarching Discussion for the entire symposium. The goal of this, it's going to be approximately one and a half hours of discussion. Basically, our goal is to be able to get something out of this symposium. To be able to think about what have we learned about AI for social good as a whole and synthesize in a very concise manner what is it that we should take away from this symposium.

I would like to start by going back to the questions that I raised at the opening talk of the symposium, the opening address. The first question is, have we arrived and we're going to be very interactive. I'm going to put forth questions. I'm going to propose some answers, and then I'll go around the room to see if other people have comments. The first question is, have we arrived at some sort of a broad understanding about what kinds of research would we like to see in this space of AI for social good? Any takers?

Speaker 2: I saw a lot of quite different types of research and actually, this gives me an opportunity right away to come back to your slides from the last session. Where you presented the model of working, where there is an expert from the field giving a problem a to computer scientist, which then provides a solution that goes back. I think, at least for me-

Amulya Yadav: That may not be the only one.

Speaker 2: The cooperation is we, together with expert, look at a problem and together we find a solution of which the computer science part is part of the solution. That gives you a completely different picture and also indicates a completely different type of research that should be done.

Amulya Yadav: Okay, fair enough. Yeah, I guess that's true. That is not the only picture that we ... That is not the only model for AI for social good research, there can be many more models. Anybody else?

Fei Fang: I'll just add a comment. Just to add a comment to that, I believe that it is a collaborative effort. It's not like we take the social good challenge as input and then try to come up with some solution. Then just output to the social good area, there will be a lot of interactions. Go back and forth discussing the model and trying to build something together. That's exactly what we did in our pause application, where we talked to the NGO, our collaborators from NGO's who has done a lot of patrols. We have to talk with them and together build a model that is implementable in practice and from there we start working on the computational challenges.

Amulya Yadav: Yeah, go ahead.

Speaker 4: Following on what Fei says, I think the answer to your question is, there is only one kind of research that we need to do in AI for social good and that is multi-disciplinary research.

Amulya Yadav: That’s true.

Speaker 4: Starting from multi-disciplinarity and not from the different disciplines. That's where I think all the comments come a bit to together.

Amulya Yadav: That's a great point. Just to play devil's advocate, what do you think about research that is not necessarily multi-disciplinary to begin with but it has potential to be extremely multi-disciplinary? For example, it may not be solving a problem for social good right away the way it is construed, but you can see, or the authors argue that … The way they argue that they show you how much potential this open-ended research has for social good. Do you think we should include that kind of research in AI for social good or is that something that we would want to keep separate from?

Speaker 4: As I see it, there is a lot of research in AI, which we need to do anyway on AI topics. We can use examples or cases from social good to do that research in AI, to develop our AI methods and tools and technologies and so on. That is to my view a bit different from the research on AI for social good. That I really think we have to start from the multi-disciplinarity in which each of the participants brings together the tools and the methods that have been developed in its own discipline. The issue is that when we are doing AI for social good, from my perspective we really have to start from the multi-disciplinarity.

Amulya Yadav: Okay.

Speaker 4: When we are doing research on AI, it's very good to use social good as a kind of a case study for our AI research.

Amulya Yadav: That’s true.

Speaker 4: Which is slightly difference but I think it's an important difference.

Amulya Yadav: It would be good if I can keep adding stuff as to what we are arriving at. All right, is there anything else that somebody would like to add to this question before we move on to the next question?

Speaker 5: For me, it was very interesting that many of you focus on animals also, not only on human beings. Most of you know this document, it's a very good document in my opinion, but they do not mention animals.

Speaker 4: If you want to join please, we have asked you for opinions so please make your opinion now.

Speaker 5: The space is defined in part by the areas of application that we are working with. Anything can be called social good right people say high frequency trading. Well, that's also social good in some way but that's not the kind of social good we really mean. It's a fuzzy boundary as to where we draw the boundary but it seems the kind of domains we would like to see in social good are ones, which typically are not of commercial interest, typically are leaning towards low research communities, towards populations that have not benefited from AI or technology more generally. That's the set of allocations we want to go towards in terms of social good.

In some sense the research then is derived from the applications, it's user inspired research that is not to say that it's application but it's derived from that space of applications. As such then there's certain characteristics that may be true of these domains, which may not be true in other domains.

Amulya Yadav: We'll get to that yeah.

Speaker 5: The research should be tailored, it's not that there's some research we do and we're necessarily looking for application but we start from the problems and look from the problem as to what research those problems require us to do.

Amulya Yadav: To add to that there are applications where there may be commercial value in them but they may still be doing social good in the sense that we want them to. An example, a case in point is there was this free basics program that was introduced by Facebook, that was going to introduced by Facebook in India. Where they were trying to provide free internet to all people in India who don't have access to internet, but and there's a but. They would regulate what gets shown.

They would only provide the most essential services like WhatsApp and Facebook of course and that is why this program fell through. However, the intent behind the program was one of social good. It would have benefited the end users, the people who did not have the internet would have internet but what was wrong with that is that they were monopolizing because it was of commercial value they monopolizing the market so that no other future player could enter this new market that they were creating.

Speaker 4: [inaudible 00:09:33] with question number one should be to understand what is social good, what is good?

Amulya Yadav: What is good, question zero.

Speaker 5: I'm saying that to add user inspired research as not just a multidisciplinary research but the research that's derived from the problem as opposed to the kind of research that may be driven from some other extensions or whatever else that people want to do.

Amulya Yadav: You're saying that it may be the case that you may have research, which was not intended with the aim of doing social good but it may have done a lot of social good for example … That is not something that we would want to include in our discussion or it's not as black and white?

Speaker 5: The question was, what kind of research? What kind of research would we like to see? I'm saying user inspired research is the kind of research we would like to see.

Speaker 6: I actually, before even showing up yesterday I was puzzled by the title social good versus public good coming from an economics background it's a term we normally use. They're seeing all the presentation there was an interesting distinction that my not have been intentional but generally when we think of public good we think of something that you can't exclude someone, it's non-excludable. What we saw today transportation, environment we saw a lot of those.

The one distinction is we didn't see a lot of national domestic security, which is generally included in the public good definition. What we essentially said is it's like a consumer focused public good is a lot of what we are, and that's where you get into low income groups. I thought that distinction was made but it was made subtly and that was interesting.

Speaker 2: Just following up on Melon's comments. User centered problem, there is in the Netherlands we had a very extensive exercise to make a national research agenda where really the public could actually also give input and a lot of input. There was like 12,000 research questions coming out of that whole exercise. That's categorized and whatever … Recently I was in a meeting about all that and what came out, it's like 90% of those questions they're inter-disciplinary, multi-disciplinary questions. Saying user centered actually means nowadays multi-disciplinary almost by definition.

Amulya Yadav: That's interesting.

Speaker 7: I have a question maybe that is there differentiation as the previous speaker? I'm sorry I've forgotten your name. I was asking the difference between the public and social good and the nuance being it seems to be a little bit more of an, as you said an economic basis or is it associated more with standards as norms. Therefore, if something is AI for the social good, if it's standards and norms then it's more AI. Hopefully, all AI is for the social good and that's more dependent upon and you mentioned before about the possibility of looking at the Belmont Report.

Actually, there's the Menlo Report, which is the follow-up from that. It might be the refining of the Menlo Report rather than the Belmont Report. Perhaps, again, the difference between public social good is an important one if you're going to be talking about what kinds of research to be in the social good.

Amulya Yadav: That's a great point.

Speaker 8: Coming from the social work research perspective, one of the things, one of our guiding principles are social justice, when we're thinking about research it's really within population spaces, groups that are historically, systematically marginalized. That's an underlying theme for us and if we're thinking about it, it's public good, social good that's one of the touch downs I always go back to. Is this a group that has been suffering from some sort of injustice, coming back to social justice?

Amulya Yadav: You think that is how we should define our problems?

Speaker 8: Those are one of the things that I think about when I'm doing research [inaudible 00:14:35].

Speaker 4: Looking at the types of research that were presented, I think it's not just user inspired research but it's research that benefits our focus on fundamental human rights and fundamental human values and all the issues around that. That's where we can distinguish it from public goods or other kinds of things. It's really about what is the fundamental human values and how we can make those values really and human rights really a right and affect for most of the population.

Speaker 9: That's essential. Unfortunately, every field has its own language and way of talking about things, but this human rights it's probably a more public or popular way of talking, a more lay if you would say a way of talking about what within social work when we say social justice. What we're really talking about is human rights and human welfare and the well-being of people and particularly as Robin was suggesting that there are ... As you're enacting solutions you don't want to further marginalize people who have been marginalized and you may even want to be thinking about projects, which are about raising up people who have not previously been. This is exactly what it is.

Amulya Yadav: That is the case maybe I mentioned the Belmont Report, Karen mentioned the Menlo Report maybe another point. A starting point for us could be the UN Human Rights Commission, the principles that they follow.

Speaker 10: I actually wanted to at least expand this idea beyond just human rights because as someone previously said that there's a lot of animal environment that goes into this. We talk about marginalizing people but it's a little more broader than that. Is there an entity of some sort that can be benefited from artificial intelligence that they couldn't do by themselves?

Amulya Yadav: What do you mean by entity in this [crosstalk 00:16:50]?

Speaker 10: Well, you have zebras that are not going to be looking after their own self-interest by themselves to maintain their own existence, you're not going to have poor people who don't have the resources and the facilities to look after their own human rights interests. It's not as narrow as just looking at.

Amulya Yadav: That's a good point but I would imagine that that definition would encompass for example the example that Melon was giving high frequency trading. The people who are trading would benefit by using AI, they would earn more money but-

Speaker 10: Maybe we just haven't been imaginative enough to find the social good in that avenue, not to say that it doesn't exist inherently in that profession, we just haven't found it yet.

Amulya Yadav: That's true.

Speaker 6: I don't know that I exactly has an opinion here, it's more of a question but there was … When I think of the traditional stuff, going back to the public good. One of the things that was maybe under represented was education, which when I look at definition of human rights and values I would include education in that but I didn't really see that represented so just thoughts.

Speaker 5: I find the public good and social good distinction very interesting and intriguing, we don't want to exclude security and policing, those are also important things to have of course. I guess it's just a very interesting point that hasn't entered this discussion.

Amulya Yadav: We've had people with security as well.

Speaker 4: Does it distinguish between what kind of [inaudible 00:18:25] case studies we present here during this today, can the type of research be done in social good and [inaudible 00:18:33], we just didn't have them to be here.

Amulya Yadav: All right, anymore?

Speaker 7: Just one more, thinking in terms of our meta-humanity and that was indicated somewhat with by including animals, the planetary. I don't mean to sound too far out there but I am. The fact is our inter-global inter-dependence socially and environmentally and our justice systems and all of those things are increasingly with the AI community impacted by the AI community at a global level. From my point of view going too big, you don't have to stay there all the time but being inclusive of that a form of radical inclusion of all of those animals included and the spiders is important to be considered.

Amulya Yadav: That's a great point.

Speaker 11: I'm more thinking of not where we can apply AI for social good but rather intrinsically can we because AI has so much penetration in how as a society has been shaped in the last few years. If there are any social bad that AI has introduced or building technology like addressing social bad, would that be AI for social good like say for example looking into eco-chambers created by social media and the impact that it has on political and other spheres. The data that most of the AI techniques are trained on might reflect the various stereotypes that they're accessed and can we neutralized that through within our techniques? Those are not necessarily about applying AI for a specific domain but building techniques that takes into account like an objective of social good intrinsically. I don't know if that counts as the AI for social good.

Amulya Yadav: I believe so, I mean.

Speaker 8: That's something that I've always thought is not only is AI for social good making new technologies but it's holding an ethical, some standard to technologies that already exist or I don't know questing ethically or whatever, what AI is out there and is it really social good and what not.

Amulya Yadav: To follow up on your point maybe if we focus on ensuring that social bad is not happening that is.

Speaker 11: Or building techniques that can actually control for that, some way of controlling or having … We talked a lot about interpretability and all that, all of that comes into that kind of question.

Speaker 5: That's an interesting line of thinking and I wanted to just add a little bit of additional context. Last summer the previous White House Office of Science and Technology Policy had four workshops on AI and I guess CCC was involved with the one on AI for social good, but they intentionally divided up the topics of impact of AI on society into four different things. One was AI for social good, here there was a lot more emphasis of actual positive applications, what we can do today positively. There's safety and control, which is more like how can we control AI from killing us or something like that. It was much more defensive if you will.

There was one on economics and jobs and things like that and then there's one more on ethics and law and so on. All of these are very, very important areas for us to think about but in some sense, I guess if you take on the whole thing that might be too much. In some the idea for this community if you will is to take one part of it and not the whole thing. I guess that would be my suggestion.

Speaker 9: This reminds me of, there was a Supreme Court case in 1964 about pornography and one of the decisions about the Supreme Court case said, "Well, at some point we could continue to try to define what pornography is but I know it when I see it." It almost seems to me like it's really important to outline some basic contours but perseverating endlessly on getting it just right may stand in the way of us moving forward with what we as a community may understand more intuitively seems like what it is.

There may be as the supreme court justice who in his decision said, “I can’t define pornography but I know it when I see it.” One of the ways that we may do this is we may say, “Well, here’s some features that we care about and then where are some examples of things that we think are instances so know it when you see it.” I just fear sometimes these … It’s a fascinating question and it’s well worth pursing but at some point, you might need to let go in order to let it breathe and not get so stuck in the definition.

Amulya Yadav: That’s a great point, what we’re arriving at is that we can have these some defining principles for what is AI for social good research but it’s impossible in a way to completely characterize the space, to completely define boundaries that everything that falls outside of this is not AI for social good research and everything that falls inside is indeed AI for social good research. As Eric was saying we can decide it on a case by case basis as to what is AI [inaudible 00:24:42] research and what is not. Should we move on to the next question?

The next question is, are there any common unifying research challenges that lie beneath most problems in AI for social good? I would like to start that off by highlighting one feature of research that, one kind of research that has been coming up again and again in many talks that we’ve seen in the last two day. That interpretability of any system that we want to deeply and I believe that it is not a feature that’s something specific to AI for social good reach. It is going to be associated with any sort of research that is going to be used by humans, any sort of decision making research that’s going to be used by humans and AI for social good research just happens to be one of those kinds of research.

One research challenge that’s common across many problem in AI for social good research, is we have to think about how can we make the system that we’re making more interpretable so that the end users are going to be more willing to accept those solutions. Because you can do all that you want to do in your lab but if the reason you built it is it should be used by somebody and if that somebody is not willing to use it then what is the point. The interpretability has started to play an important role and moving forward it is going to play an even more important role, any sort of comments on that.

Speaker 12: Two comments, one to follow up on what you just said. Based on work that we’ve done with the police department in Nashville. They really don’t care about interpretation of our models or for example they’re not exactly interested in the math that goes behind the model. Interpretability comes into the scene based on what the consequences of not interpreting a model could be. There are so many applications, which are extremely result driven and what the agency trying to deploy it is concerned with is whether it can give good results or not. That is in response to what you just said.

My view on what the common unifying research challenge is-is making things and data open source and Fei's talk, when she was talking about what kind of researchers should come together to work on AI for social good. I'm not referring to a particular type of researcher but I think we all realize that governments and cities need to come together. For example, most of the work that I do with the police department in Nashville has this dataset that is confidential.

For example, I collaborated with your lab but then in order for you to work on the Nashville dataset you have to collaborate with me and something like that. It's not available for anybody to use but you do get the point that if the data is perfectly anonymized anybody should be able to use it. More and more as people really score as well as data a lot of research challenges would get addressed by more people working on the same data also giving people the same platform to work and compare their methodologies on.

Amulya Yadav: That's an interesting point, access to data is how to get access to data. Although I'm curious why is interpretability not a big issue with policemen, why would they not care as-

Speaker 12: The work that we do, we don't exactly ask police whether or not to arrest somebody or not. This is about responding to calls as fast as possible as long as they are happy with the results they don't exactly care about what happens under the hood. Again, this is a very specific scenario in Nashville, I'm not sure about how other police departments take it. Again, this is also as I said it's about the consequences. If I was telling the police who to arrest and who not to arrest they would have wanted interpretability.

Amulya Yadav: I would argue that once they see the results and they see it's working very well then yes, they would not care as much about the interpretability of result but at the beginning when you-

Speaker 12: It is extremely results driven. If they're happy with the results I don't think they would take care.

Amulya Yadav: At the beginning when you would take this project to them, that, "I have this machine learning model," they haven't seen the results, they don't know how it's going to perform in the real-world.

Speaker 12: They haven't for some reason, they were not particularly interested in the interpretability but again that's one agency or department that I'm talking about.

Amulya Yadav: That's interesting, thank you.

Speaker 13: I wanted to follow-up on the comment about open access data, this is something that I've been thinking about a fair amount in the context of the work that I talked about earlier. An important part of this is the incentives that people face to make their data open or not. We all like to have open data but it may not be the case that people want to make their data open for various reasons. Scientists may worry that they're going to get scooped or that making their data available only helps their colleagues but it doesn't help them at all, why take the time to do it.

One of the things that we need to think about is how can we create systems or institutions that have, make the incentives available for people to do things like share data. For example, how could we track the use of data that other people are using so that we can give credit to the people who've made that data available. That's not a form of scientific contribution, which is currently recognize very much. Right now, it's mainly just publications, are there ways that we can expand the scope of what scientists are rewarded for so that we can have the things like open data that we want.

Amulya Yadav: How do you feel incentives basically for people to be actually willing to release their data. There's another question about open sourcing of research in general that we'll come on to next and yeah.

Speaker 14: I guess going back to your questions about verifying the solutions of the models, I guess one of the important questions, how do you validate the models, right? Typically, solutions are hard to validate in real life. You can validate them experimentally but versus how does it work and does it work well in practice and how do you validate this. I know it takes a while to do it a few years just to gather the data and a few years to test whether your solution would work or not. That's one of the problem-

Amulya Yadav: That is true, that's a great point. Eric and us and we've faced this challenge in our research, the study that we're conducting. Our estimate is that it's going to take one and half to two years and it's a longtime to do for one paper to come out. Is there enough incentive that the computer science community as a whole provides to people who are willing to do this thing? This is something that [inaudible 00:32:13] was mentioning in the morning. Another thing that ... Yeah go ahead.

Speaker 6: On your interpretability comment, there was several things in there. One of them that I heard really strongly throughout the last two days is sophistication of the user. When we were talking about social workers and we talked about [inaudible 00:32:34] centrality there's a norm within that industry that the risk she was up against. Then we talk about policing departments. Well, a politician department is going to be a much more sophisticated user who's able to understand, "I may not be able to interpret." The norm and how broadly it needs to be deployed and how close that person is going to be to the actual research. The police department probably much closer to the research and social work probably further from the research.

Speaker 8: I would argue for the interpretability, it's the role of the computer scientist to advocate that the users should want to be able to interpret it. For the police department saying, "Oh no I trust it." I would tell them, "You should want to interpret it," and that we talked a lot about the user in healthcare system with the doctor just giving them whatever. As a role as healthcare advocate we tell the patient, "You should question your doctor." Be an advocate for your user in a sense. I don't think it's okay to be like, "Oh if they don't want to interpret it that's fine," and move on. They should want, it's our role to tell them to do that.

Amulya Yadav: It would be interesting to find out that the percentage of applications where interpretability would not matter as much. The proportion of such applications would shrink as time goes on.

Speaker 4: I have a few things to add, one is participatory design. We don't want to design for the user but we want to design with the user. The example I always give to my students as well is I was a scout for many years and as a scout you always know one of the things scouts do is try to cross old ladies across the roads when the old ladies don't want to cross the road and that's something we really have to take care of.

On interpretability, one way to deal with it is being transparent and I mean not only transparent about what the algorithms do but especially transparent about how we are doing and developing the process and explaining what we are doing. Expandability and transparency is more than just making the results and the algorithms explainable. It's also explaining ourselves and the process that we take. Of course, I don't want to repeat it again but in our things the art principles, the accountability, the responsibility and transparency is important there.

What I think it's also very important is that we are doing AI for social good and that should be the center of it and not so much what kind of papers and publication we want to get off it, but we really, we should do what we preach. It's about doing it for the social good and of course that papers come out of it, it's nice but it shouldn't be the main point. It should be really much more focusing on the results for the people we are working with on doing it.

Amulya Yadav: That's a great point.

Speaker 5: I wanted to add, I mean going back to the question of validation in the real world. This is an important feature in this area that we want to intervene on something in the real world, we want to test things and sometimes it's not easy to write papers, you talked about not writing papers, but on these kinds of experiments because it's hard to get, it's far better it seems if you write a couple more theorems for the paper than to do more experimental work.

I just wanted to advertise that JR now has a special track on AI in society and in this special track specifically these kinds of papers where we do validation work, where we actually just the paper is all about just measuring things in society are welcome. This is a new track and I hope people will be able to submit to this track because it’s exactly addressing some of these challenges. It’s AI in Society.

Speaker 2: I just want to get back to the interpretability again. Also, in healthcare this happened to me as well. I got some health problems, I got some medicines prescribed and then you think, "Well that’s from the doctor, it should be good." Of course, you have to have that, the reason what it does but not exactly why this combination would be good. I encounter other people, other doctors happened to be and he’s, “Well, that’s maybe not really necessarily.” The next time I go back I say, “I don’t take that medicine anymore," and I think now we’re going to get a discussion because he gave it for some reason and he says, “Okay.”

It means that you actually should want to have this interpretability that I should know exactly why things are given also in healthcare. Apparently, it doesn’t happen like that at all. Our idea that there is a kind of complete model of things and that I do things because of that model is not true at all and I have very partial knowledge and based on that knowledge I do the best I can but it’s always partial. For us that should be also a lesson on interpretability is like you can’t do things on incomplete knowledge, not everything has to be validated as well. As long as you indicate that this is the best you can do.

I have a lot of discussion with my students especially in social simulation applications, where validation is a real issue if I do a lot of social simulation. What is the value of the results? I can compare it with past results but of course I can just tweak it in a way that it actually reflects past results and you get the same out of the simulation. Still it doesn’t say anything about predictive value. Validation in this kind of context for me is also not so clear as like what I use in physics and that’s probably worth another discussion. What do we mean with validation, what’s the value of validation in these kinds of areas?

Speaker 15: In addition, weighing in on the writing papers versus doing really good in society. There’s an intermediate point we can aspire to, which is really training people to be able to think intelligently and to grapple with these problems. The aim in writing the papers is to just get people started along this and training people in the mixture of methodologies that we hope. Hopefully, a great outcome would be other masters or PhD programs in this are to create this trained population of people who have the tools to deal with the problems.

Amulya Yadav: It’s a good point.

Speaker 9: One of the things that I heard repeatedly over the last couple of days was worrying about whether or not by pursuing a challenge that is driven by a social problem of some kind, are you going to actually be able to be doing inventive, rigorous, new computer science? I think that you know I’ve heard from you and others that you feel that often times the parameters of a real-world problem usually open up all new aspects to this. It seems that it's something that this group is concerned with to a certain extent. Happy when they see another problem that something that they can tackle that hasn't been tacked before but that seems to scratch at the back of your brain that, "Wait is this problem going to be something that there's a really new computer science problem to do or is this really just going to be an application that in another discipline might provide them with something making novel research-wise but for you it's out of the box."

Amulya Yadav: That's true.

Speaker 8: This is my last comment, just going off the healthcare stuff. There's this theory the generalization of expertise, which is applied to healthcare things but it's the idea that people that may … Coming to transdisciplinary space, sitting in a table with a bunch of computer scientist and AI experts, people that are not in that seat might assume that everyone at the table is just an expert just by … There's these assumptions that computer scientists they know all the answers and will just trust. It's this idea of, again, advocating and the other people at the table understanding that this is really a transparent experience and everyone at the tale is an expert in some way and yeah, I don't know it's a good article to read Generalization of Expertise.

Amulya Yadav: To add to the discussion on interpretability, one thing that we focus this discussion on is interpretability for the end user. For example, in the doctor's case is the prescription, is it interpretable for the end patient. There's also a case to be made for interpretability for the doctor himself because many times we are in a situation where human beings don't want to relinquish control and it may be a very good idea for them not to relinquish control.

In such a situation is it the case that we always whenever we're thinking about these problem, do we want to think about designing agents or algorithms, which don't replace human beings per se, but instead work as a team? Only provide suggestions, only act as a decision support system as opposed to just replacing them?

Another unifying research challenge is that since most of this research is going to be used by human beings, we again get into the issues of protocol surrounding what is it, what are the protocols that should be followed when our research interacts with human beings and that is the things that happens in IRB? Those protocols have become very important and to that end do you think going back to the guiding principles make sense? What should be the principles that guide our development of research when it comes to interactions with human beings?

I outlined three principles, the benefits and respective persons. Do you think they are the right principles to be followed or do we need a completely different set of principles because we are in a different domain, those reports were created with a different mindset, times have changed?

Speaker 4: Similar but a little bit different set off principles is ones using in medical research, which are principles of beneficence, nonmaleficence, justice and fairness, which in a lot of aspects also fit very well the type of research that we are doing.

Amulya Yadav: What beneficence and nonmaleficence are the same thing right?

Speaker 4: No, because you can choose between doing good or not doing bad to someone. It’s a different way of looking at it?

Amulya Yadav: What were the others?

Speaker 4: Justice and fairness.

Amulya Yadav: Does anybody have to add to that?

Speaker 13: I’d just like to add that maybe a prior question is even should we be the ones coming up with the guiding principles? For example, in an earlier talk someone asked, well should our algorithms have a utilitarian ethics or Kantian ethics or Aristotelian ethics? I asked myself if I pick an AI expert and ask them what they think versus picking a random person off the street, I’m I likely to get a better response from the AI expert and personally I’m not sure.

I also worry that part of what’s fueling the latest outburst of populism around the world is a feeling that experts are making decision on behalf of people in a way that’s not transparent to them or which they haven’t had adequate say in. I wouldn’t be here if I didn’t think that AI for social good could have a lot of value but one of my worries is that it comes to be seen as something, which is just another way in which the system is not accountable to people and the decisions are being made in a very black box way.

For instance, like Facebook’s news algorithm. That has a very big impact on the world but by what process is that being determined? Those kinds of scenarios will only become more and more frequent as AI becomes a larger part of our lives. We also need to be thinking about not just what are our principles but what gives the legitimacy and what justifies them in the eyes of the public.

Amulya Yadav: To add to that another thing that would become very important as we move ahead is as humans, as we have more and more algorithms that start doing the work of humans. When you have human beings who make mistakes we know who made mistakes, but when algorithms make mistakes who are the ones that who are actually responsible for that but that’s a separate discussion to be had?

Speaker 16: How do you define algorithm making a mistakes?

Amulya Yadav: I guess for example we had [inaudible 00:47:01] there it's quite obviously what … Any machine learning algorithm when you have a fast positive or false negative that's a mistake, right? In machine learning applications and classification problem it's very simple to identify what's a mistake and what's not a mistake and what is the consequence of that mistake and who is responsible for that mistake? Those are important questions.

Speaker 16: I was thinking of a more personal error or implementation error.

Amulya Yadav: That is a great point, there is … Does everybody know of FMRI research? There was this technology called FMRI, which is magnetic resonance imaging high frequency. They built the software and there was 10 years of research, 10 years of papers being published using that technology and after 10 years people realized that there was a bug in the code. 10 years of research completely down the drain. I guess that's an excellent question, what happens if that happens, who's responsible and what do you do to the person who's responsible, he made a mistake. This came out last year, this was reviewed last year 10 years everything has been … There were conferences, there were separate conference just for FMRI research.

Speaker 16: All of them [inaudible 00:48:26]?

Amulya Yadav: Everything on.

Speaker 7: I was just going to say, I'm not sure that we're going to be able to come up with the guiding principles at the end of this. That's quite a large conversation that would deserve quite a bit of time of its own. The only other thing would just be to add in there, partly it has been said. Just that there is a tendency for Ai from the average person perspective to be very much a black box and so mysterious, unknowable. The average when talking about AI, not the average person in this room, okay. The average person out on the street.

Ai, it's a mystery, it's something they don't know anything about and you begin to explain or give an explanation, it's already gone. This came up in a number of places, trust is huge. It's going to require a great effort, I believe, from the AI community to ensure that, to go the extra mile to ensure that what is being done by AI is explained. That's just an extra urging this community to go that extra mile because it can come back and bite us.

Speaker 17: I'm not sure if my question is, actually falls in the guideline principle but being a researcher, whenever I'm doing a research there's something that always bugs me. Whenever I find out data on web I know that data is actually from the people who are very active on web, who are very opinionated or who actually express themselves on social media or web but what about people who are not that opinionated or who don't speak up? It's like whenever we do any kind of research on data, which is available on web or social media we are talking about the sample space who are very active.

Amulya Yadav: Which is very biased.

Speaker 17: Whether we accept it or not that we are biased on whatever paper, whatever AI, whatever classification system, whatever model we bring and we deploy they are based on those very small sample space. I'm not sure where I'm going with this but this is something which always bugs when I'm doing my research.

Amulya Yadav: That's a great sort of common research challenge, this challenge and this question of data being biased. You only get data from where you get observations this shows up in Ben's research in wildlife protection. You only get observations about wildlife poaching from where the patrols have gone, right? This showed up in Jason Stoke also. You are only going to get pictures of animals where the tourist van takes them. That's a great unifying research challenge.

Speaker 18: [inaudible 00:51:28].

Amulya Yadav: Data bias.

Speaker 5: Going back to the point made, which I agree with. Is it really the AI researchers who should come up with the principles, I think that's a really excellent point. On the other hand, I guess if we just leave it to the ethicists or whoever else they may not know enough about AI to really understand what questions to ask, what limits to set? There needs to be some collaboration in order to make this happen.

This is also the principle in the AI 100 Report that was put, that was generated, where there were economists and others who were on the board so that they all participated. The sections of the report, which deal with economic aspects and so forth where it's a collaboration between AI and the particular economist or whoever else but I agree that this is not just our jobs, it's a collective responsibility.

Amulya Yadav: Any other thoughts? What time are we wrapping up?

Fei Fang: [inaudible 00:52:57].

Amulya Yadav: 15 more minutes okay. I guess if that is the case let's move on to the last question and then if we have more time we can come back. This is a brain storming question, what research would we like to see? We have seen a lot of research happen, being talked about in the last two days. What research do you think has not been done yet, what research should we be focusing on, what problem area should we be focusing on, what are the problem areas that need our help? By our I mean the help of AI the most.

Speaker 17: Long back I read a blog about how social media AI is actually making people depressed and it's changing their behavior. It's like if a person who's not married she constantly sees a lot of people getting married she gets depressed, people who are not outgoing and maybe if I'm not connected to anyone and I'm content with my life. If I'm not going out, if I'm just sleeping my way out on the weekends I'm okay with it but as soon as I see that my friends are going out, I feel the need that I have to go out.

I see even there are videos and there are many movies, there are many shows which are coming out where people show that how other people's life and just because you think they're happy about it you try to change who you are. At the same time you see their life they're just going somewhere taking pictures and coming back, they're actually not enjoying.

People are actually taking pictures, creating memories which doesn't exist. The first-time reason all this social media, all this way of being connected is actually making humans depressed. If somebody actually has to do something AI for social good, some way to actually make us happy, actually happy in a way we actually care about, not happy in a way that people think we are happy.

Amulya Yadav: AI used to cure depression on social media, AI don't make us happy. Anything else?

Speaker 19: I just wanted to add something, in the beginning of February I was at the Triple I Conference and I noticed that there some of the sessions were stiff divided along classical divisions such as Knowledge representation and reasoning or machine learning or this and that. I'm a Knowledge representation and reasoning person, you probably have noticed that. I would like to see and we're talking about inter-disciplinary research between artificial intelligence and other fields such as social work or medicine etcetera. I would like to see a lot more collaboration between areas of artificial intelligence.

Amulya Yadav: Between [inaudible 00:55:55], okay.

Speaker 19: Yes, personally I would love to collaborate with machine learning specialist because what I'm building can benefit from machine learning and I feel the problems that they are trying to solve would benefit from our expertise in the field of decision support systems.

Amulya Yadav: That's a good point.

Speaker 7: One of the points similar to that-that Carla made in her opening remarks that what is very valuable is to have very diverse perspectives engaged on issues because that broadens our thinking and our minds.

Amulya Yadav: Anyone else?

Speaker 9: I mentioned this earlier but I think that there's a flip side to this bias and the garbage in, garbage out fear about biased that has been talked about a few times today. I think from an AI for social good perspective. If we can do more research that is intentionally trying to uncover biases in the systems and in the data, which is used by machine learning or other approaches. That gives us a way of actually going back to those systems and essentially saying and this is maybe thinking about Venard's natural language processing work, right?

You find out that in fact police are, and it's not a surprise, engaging with black motorists differently than they're engaging with white motorists but there are policies at the institutional level are to not engage in racism and yet you can find these very micro level … Aggregate these micro-level instances of discrimination and bias that are really not just at thing to say, "Oh gosh our algorithm doesn't work," but this is actually of social value to have these things uncovered form these systems. I'd love to see a whole bunch of trying to create or trying to uncover I suppose bias not trying to create it.

Amulya Yadav: That's interested, I read a paper in which there was a machine learning system and they were trying to do some prediction on who's going to likely to commit a crime and who's not likely to commit a crime. They had racism as a feature, they didn't want to use racism. They removed that race feature from the dataset and then trained the model but it turns out that their model was still being racist, and why was that the case?

Because they were using an address feature and the address there was a correlation between areas of the city. Different areas had different proportions of populations living. The poorer populations invariably tended to be of black and that is why it was still being racist when they tried all they could do to not be racist. That's a great point, how do you uncover this bias despite you're trying to make your algorithm not racist, to not have any bias towards a particular race. You will still end up in a situation where you still have a bias.

Speaker 9: The difference in what I'm saying and what you just said is that there is value to society in being able to say that this machine learning algorithm, which is based on data which is from existing systems. Thinking about the bail prediction algorithm for example. If you find out that the bail predication algorithm is in fact biased systematically against a particular racial group because all of the judge's decisions that you're modeling it from essentially are biased against a particular racial group.

This is in and of itself evidence of bias that exists in those systems that is largely being glossed over because policies and laws have been that say, "You cannot or should not act in this way." Because those laws and policies are in place systems think that they're not acting in this way because they've said, "Oh no, no, no we have a policy, you can't decide to put somebody not back on the streets or not based on their race. We don't do that," and yet you can find through these algorithms that people are doing that.

It's almost like a social advocacy aspect of it. It's like when these algorithms are turning out to be not desirable in your case, it may actually be evidence that could be used in a different context. It’s not necessarily just, “Oh you’re done, this is garbage, move on to the next one to try to make it better.” This may actually be evidence for social problems.

Amulya Yadav: That’s right yeah. I guess Fei, wrapping up. Yeah.

Speaker 2: I’m going to advocate my own type of research. There should be some more conceptual research especially in this area, where there’s very multidisciplinary research. Sometimes you need to step up one level of obstruction in order to be able to connect to the kind of concepts that are used in other disciplines.

Amulya Yadav: Okay, I see.

Speaker 2: That’s quite complex and also not very rewarding in the short term but it’s very rewarding in the long run.

Amulya Yadav: Okay, fair enough. This is too make sure.

Speaker 2: When you get to concepts like failures and all that kind of stuff, it’s really the social concepts that are going across-

Amulya Yadav: That map across different disciplines. We have arrived at the end of The Overarching Discussion. I hope we all learned something out of it and with this we end. There’s going to be the plenary session talk in which Fei and I will summarize the entire happening of the symposium in five minutes.

Speaker 20: I’ve heard it’s going to be funny.

Amulya Yadav: That is the intention, that’s what it's supposed to be. Let’s see if we are able to make it funny. I would like to thank you all for being a part of this symposium, it’s been a great learning experience for me personally and thank you.