



Reflections on AI

Q&A with
Y-Lan Boureau

“The biggest promise of AI in wellness is scalability.”

The TUM IEAI had the pleasure of speaking with Y-Lan Boureau prior to her 24 November 2020 Speaker Series Session on *Balancing issues of trust, helpfulness and responsibility in wellness and conversational AI*. We were able to ask her some brief questions about AI ethics, the role of academia and research institutions in creating AI frameworks and the ethical challenges arising from creating conversational AI and nudging people to live a better life.

1. What is the biggest misconception about AI?

One that strikes me often is that people get a sense that the challenges of AI are very unique, whereas I view the challenges more like an amplification of existing challenges. There are questions that we have been grappling with for a very long time, and they get amplified by AI. Because with AI everything is bigger and faster, it puts more light on the existing challenges that were always there, but perhaps less apparent.

2. What is the most important question in AI ethics right now?

Actually, there are two that strike me. One is the sense that communicating to people often is not a technical question. The big question is to define what is good. If there is a right thing to do, decide what it is. Often people disagree on what is the right thing to do. It is not black and white. There are complicated tradeoffs and, unfortunately, in a complex world and in complex

situations, people are competing in a way that is not easy to understand, in particular for big statistical phenomena. Think of COVID-19 for example. People have an idea of what is the right thing to do, they often think in terms of representative narratives and do not see the bigger statistical picture. For example, at the very beginning, when you had exponential contamination, it was mathematically certain what was going to happen, but a lot of people were not getting that sense. The big challenge is how to communicate complicated ideas that sometimes don't fit into one sentence.

3. Who should be in charge or involved in developing ethical frameworks and standards for AI?

I think the most important thing here is that it should be a collaborative framework where many professionals from different disciplines come together because they all have different perspectives. You need to have people who will understand the techniques themselves, people who have a philosophical or ethical background, e.g. sociologists or psychologists. It needs to be interdisciplinary because they are going to understand different parts of the problem. As I mentioned before, it is really a tradeoff. You need to understand all the sides of the tradeoff before making a decision about what it is good, what it is desirable, and what it is problematic.

4. What is the role of academia, academic institutions and other research centers when it comes to the ethics and governance of AI?

I think that the role of academia is both to provide the expertise, the studies and analyses of what is happening and also this sense of permanence of studying and trying to understand and giving this long-term view of the tradeoffs, both for good and for dangers.

5. How can the use of wellness and conversational AI systems impact our everyday lives or the way we live?

For me the biggest promise of AI is scalability, i.e. that a lot of people who do not have access to resources can get a better system that can be personalized to them at very low costs or even for free thanks to the systems being automated. For example, if you think of wellness and trying to find somebody who can help you learn wellness techniques or meditation or these kinds of things, then with conversational AI you could have your own personal coach that could talk to you. So for all those people who are not able to afford these resources, thanks to scalability, we can bring that to them through AI. The second aspect of that is personalization. Given that a lot of AI is based on learning, the systems can really learn from the everyday interaction with the users. A lot of research has shown that people actually feel more comfortable sharing things with a system that won't judge them because it is not a human, so they feel they can be more open without any fear of stigma or judgment.

6. What is the biggest challenge in the development and use of ethical wellness and conversational AI? How do we balance these, sometimes competing, ethical challenges?

For me, the biggest challenges are just the same as in regular life, i.e. how to help people. There is this tension between the

different values that we have: between freedom of choice, so control of your own personal choices, and doing what is good for you. For example, if some people really want to eat at McDonald's every day, it is their choice. However, it is bad for them. Should I force them not to do that? And what about somebody who never wants to exercise – like, for example, my dad. I have a choice between forcing him to exercise and letting him choose for himself even if he is choosing wrong. He might sometimes not even have the tools to realize it because he does not want to hear about this. So, it is this tension between several values, the sense of freedom and the sense of what is good for you and what does it mean to help someone? We see a lot of that here in the US with the COVID-19 crisis. There are people who get really distressed by the idea that they have to wear a mask and that this is mandatory in some places. These people feel that this is a big infringement on their personal freedom. On the other hand, you have public health and the scientific consensus that it is actually very helpful to wear a mask. So again, you have this tension of values. For wellness there are just the same challenges, the ethical challenges of what does it mean to help someone? Where do you draw the line between personal freedom? And what can you do for people as opposed to what they want and what they decide for themselves, knowing that for everyone this is very different?

7. Why is the interdisciplinary approach so important when it comes to AI Ethics?

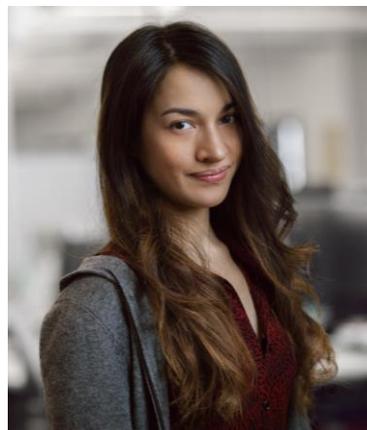
For me, a society functions best when people do what they are best at and what they are knowledgeable in. Interdisciplinary means that you are going to take people from many different places. And if you think of the ethical challenges,

there are many dimensions that you have to get right. There are many tradeoffs because you have benefits and potential risks, and you have to weight them. You need people who have a sense of the history and of what has been done in the past. You need people who understand these systems we are talking about. You need people who get a good sense of what the future might look like, and you need people who have thought about the social repercussions of things. So, you really need to bring people to the table who have all these different angles.

8. We often say that AI is changing or transforming the world. To what extent is AI changing us as humans?

I think, again, one of the important aspects here is scalability. The fact that AI can amplify and magnify things. Often, it is things that are already there, but they are just bigger because you can get a big scale and amplification with AI. It is like a bigger tool and everything gets bigger. AI forces us to find ways to adapt to a world that is moving fast. What we see around us is not our environment so much anymore. What I mean by that is that we often react to things with narratives we can understand easily. This sense comes from experience. But now with AI, we can connect with people on the other side of the world. We can get connected and share the experience with people who are really far away because everything is so scalable. Many AI models are trained with data from a lot of people. This means that we get connected to a much bigger universe than what we are experiencing day to day physically. This forces us to change a bit how we think about our intuitions, what we know to be true, and how we interact with the world.

Meet the expert



Y-Lan Boureau is a research scientist at Facebook Artificial Intelligence Research, where she focuses on building more helpful conversations and understanding dialogue. She received her PhD from New York University and École Normale Supérieure (within the INRIA Willow project team), working in machine learning and computer vision, under Yann LeCun and Jean Ponce's supervision. She went on to do postdoctoral research in experimental psychology and neuroscience at New York University working with Nathaniel Daw, investigating self-control and meta decision-making. Her research strives to foster stronger people orientation in AI. She is one of the directors of the ELLIS program on Natural Intelligence.