



Reflections on AI

Q&A with
John Tasioulas

“Philosophers are well placed to highlight ethical questions, which are subject to rational investigation, discussion and debate”

The TUM IEAI had the pleasure of speaking with John Tasioulas, professor of ethics and legal philosophy and director of the Institute for Ethics in AI at the University of Oxford .

1. What is the biggest misconception about Artificial Intelligence?

I think there are many misconceptions around AI, but perhaps the most fundamental one is that it is, in some sense, an inexorable technological development which we, as individuals or society, can exert very little control over. However, that is manifestly false because it is a terrain where human choices are constantly being made [...]. Of course, it serves those who do make these choices, those who are currently in power in the existing status quo, to disguise that fact by presenting them as unalterable facts of nature. I think that is a misconception we have to constantly fight against and highlight the fact that this is the domain of choice whose choices are being made here, which values really important questions we have to constantly stress.

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

I think it is very difficult to choose one specific question as the most important, so I am going to cheat a little bit and say that there is a kind of general two-part question that is important and that should frame a lot of research in this area. One is the question of which the values are that we want AI to further. Secondly, what is the

best institutional mechanism to ensure that it does so with those values. So in the question of values, I think there are some serious misconceptions, and in particular, there is a tendency to identify human values with actual preferences of people. And that to me is a deep mistake because preferences are just facts about people. These facts, these preferences, could be based on false information, they could reflect varied prejudices (racist, sexist and so forth). The point of ethics is to subject our preferences to critical scrutiny. And on the institutional front, I think the really important question is how do we secure some kind of genuine democratic control over AI developments. I think that is really crucial. And, on a more positive front, how might it be possible to use AI to enrich our democratic culture? For example, to enable more radical forms of citizen participation amongst the citizens, who often feel alienated by AI developments, and feel that their voice isn't heard.

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

For a long time there was this misconception that somehow big tech corporations would be at the forefront of developing ideas about AI ethics and governance. And clearly that idea is a non-starter, even before the recent scandals affecting in-house ethics in big tech corporations. I think that the driver for thought on AI ethics and governance has to be a vibrant civil society that is inclusive

and engages informed debate. Universities have a very important role to play, partly because of their academic independence (although that is a precious thing that is constantly under threat and something we constantly have to fight to preserve) and partly because universities have the capability to facilitate the kind of multidisciplinary discussion that is absolutely necessary in this area. But above all, what I want to say is that academics have a special obligation to model civil and rational debate for a wide citizenry about contentious ethical questions and if we can't engage in such debate, then it seems to me - given that we are in such privileged circumstances - there is little hope for people outside the university sphere to engage in such debate.

4. What are the most important things that the COVID 19 pandemic has taught us about the connection between AI and human rights?

There is a temptation for everyone to read their own political prejudices into the Covid pandemic and into the often inept response of governments to that pandemic. What I would say in a very general level, is that the pandemic illustrates or casts into sharp relief a certain question and that question is: 'In the service of which values and which interests is AI being developed?'. Is it being developed in the service of profit-making and often in a way that involves manipulating or exploiting people's preferences or beliefs, often in ways that perhaps violate their human right to privacy or violate their right to political participation? Or is it being developed in a way that, for example, serves basic human interests and rights? For example, the right to be in a safe working environment or the right to access to life-saving vaccines. I think we have a serious problem here. The

answers to those questions are not particularly encouraging at the moment.

5. How can philosophy help us when formulating governance or policy approaches to AI?

Philosophy has an important role to play. Philosophers cannot resolve these questions, we don't have the political authority to do so and contrary to Plato ['s beliefs], we should not be given the political authority to do so. But I think that philosophy is well placed to highlight two things at least. The first thing philosophers are well placed to highlight is that these ethical questions are subject to rational investigation, discussion and debate. It is not simply a domain for PR manipulation or simply a domain for power plays, but we can have a rational debate about value questions. That is perhaps the most fundamental thing. Another thing, I think philosophers are well placed to do, is to highlight the fact that these ethical issues do not reduce to one fundamental question. For example a question of trustworthiness or a question of human rights, however there are many different ethical considerations that often conflict, so I think that philosophy can do a great service of highlighting that fact and getting us away from some straight-jacket conception that there is one master concept that solves all these problems.

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

That is a very difficult question. No doubt that [AI] is changing us and will change us. It is very hard to predict how. So instead of predicting, I'll just simply express the fear that I have - my deepest fear about AI is that it will be corrosive of human dignity. Our dignity as human beings, what makes us distinctively valuable, is our capacity

both as individuals and as groups, to engage in rational self-determination, to be sensitive to rational considerations for and against various courses of conduct and to make a choice in light of the rational assessment. My worry is that AI might become harnessed to a wider tendency, a tendency encouraged by capitalism to turn us all into passive consumers who in a sense delegate rational decision-making to autonomous systems.

Meet the Expert

John Tasioulas is Director of the new Institute for Ethics in AI at University of Oxford which was announced in June 2019 following a donation from Stephen A. Schwarzman and launched in February 2021. John Tasioulas is also Professor of Ethics and Legal Philosophy at the faculty of Philosophy, Oxford University and Senior Research Fellow at Balliol College.

John Tasioulas received degrees in philosophy and law from the University of Melbourne and studied as a Rhodes Scholar at the University of Oxford where he completed a D.Phil on moral relativism under the supervision of Joseph Raz.

He is also a Distinguished Research Fellow of the Oxford Uehiro Centre and Emeritus Fellow of Corpus Christi College, Oxford. He has held visiting appointments at the Australian National University, Harvard University, the University of Chicago, the University of Notre Dame, and the University of Melbourne, and acted as a consultant on human rights to the World Bank.

Tasioulas works in moral, legal and political philosophy. He has advanced a version of the communicative theory of punishment,

according to which the overarching point of punishment is the communication of censure to wrong-doers. His version of the theory is distinctive in making room for the value of mercy alongside that of retributive justice.

In the philosophy of human rights, John Tasioulas has argued for an orthodox understanding of such rights, according to which they are moral rights possessed by all human beings simply in virtue of their humanity. This contrasts with a more recent view that characterizes human rights in terms of some political roles, such as being triggers for international intervention or benchmarks of internal legitimacy. According to Tasioulas, human rights have a foundation both in a plurality of human interests and in equal human dignity. Among other writings in this area, Tasioulas is the author of two reports on minimum core obligations, and their bearing on the human right to health, for the World Bank.

He has written on a range of other topics including Moral Relativism, Games and Play, the Ethics of Robots and Artificial Intelligence and the Philosophy of International Law.

