Reflections on AI

Q&A with
Prof. Dr. Huw Price

"The effort of making sure technologies are beneficial needs to be a globally collaborative one."

The **TUM IEAI** had the pleasure of speaking with Huw Price prior to his Speaker Series Session on 23 July 2020 on *The Future of Artificial Intelligence: Academia's Role in Getting in Right.*
We were able to ask him some brief questions about AI ethics, the role of academia and research institutions in creating AI use frameworks, and the transformative implications of AI.

1. What is the biggest misconception about AI?

That is a tough question. One of the reasons it is tough is that the biggest misconceptions are often only visible in hindsight. But I think it is a pretty safe bet that the tendency to anthropomorphize AI, to think that it is going to be like us, only somehow sort of more so, is likely to be a big misconception. I think it is much more likely to be very unlike us and that will make it much harder to understand. Because you won't be able to get to an understanding by extension from our understanding of ourselves.

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

Well, I was tempted to cheat here and say "your next question": the one about who should be in charge and involved in developing ethical frameworks. However, if I am not allowed that answer, then I will offer two others. One is the control problem and the issue of AI safety in the long run. However, I think perhaps in the shorter term, the issue of machine consciousness is one to be very concerned about. We need to be very sure we are not building mechanical slaves or machines capable of suffering.

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

That is a great question and I don't have a simple answer, but I do think that it needs to be inclusive in several directions. For example, the process needs to involve end users in some way, especially disadvantaged end users or representatives. It needs to be very well connected in a global sense. Because wherever they develop, these are going to be technologies with global impacts. The responsibility of ensuring that they are used well, therefore, falls on everybody's shoulders. The effort of making sure technologies are beneficial needs to be, among other things, a globally collaborative one.

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

I think that academia has several distinctive features to give it a key role to play. For example: some degree of independence from government and technology firms, access to a very broad range of discipline review points and the infrastructure for creating global communities of experts.

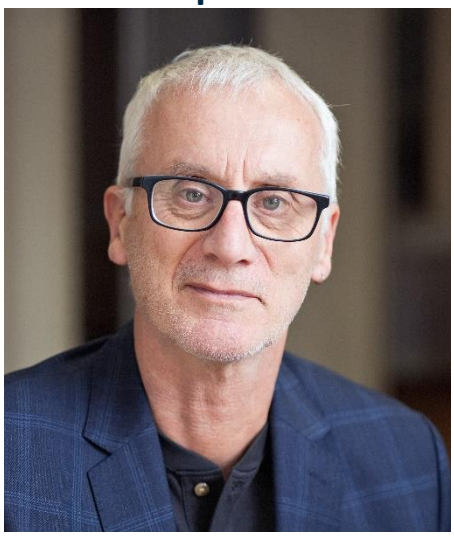
5. We often say that AI is changing or transforming the world. To what extent is AI changing us as humans?

This is the kind of question that is especially difficult to answer, except in hindsight. The biggest changes to us coming from AI are certainly still in the future. Some people seem to expect some kind of sudden step change, either for better or for worse. On the pessimistic side, for example, Yuval Harari has talked about the dangers of what he calls in his pseudo-equation *B x C x D = HH* (biology and brain science x computing x data = human hacking), the ability to change ourselves in radical and possibly dangerous ways. I am somewhere in the middle between optimism and pessimism on this. On the particular point of human hacking, I think that we have been hacking ourselves for a long time. Think of education for example or various social systems, or, going further back, thinking of the development of language. So, I think Harari is right that we are going to change ourselves pretty dramatically. But I think that is inevitable and isn't necessarily a cause for pessimism. It is a cause for caution - he is absolutely right about that.

6. With the use of tracing apps to fight COVID-19, how do we balance the ethical issues of privacy and public health?

As some of my Cambridge colleagues have pointed out in an excellent recent piece, one of the difficulties of the present situation is that we have to make these ethical choices on the run. But there are some obvious approaches for mitigating potential harms. For example, we can build sunset clauses into emergency legislation, we can assure that proper review happens afterwards to learn the lessons for future cases, and I think it is particularly important to take extra steps towards trust by establishing penalties for misuse of data beyond the needs of contact tracing apps.

Meet the speaker



Huw Price is Bertrand Russell Professor of Philosophy and a Fellow of Trinity College at the University of Cambridge. He is Academic Director of the Leverhulme Centre for the Future of Intelligence, and was co-founder with Martin Rees and Jaan Tallinn of the Centre for the Study of Existential Risk. In 2019, he joined the inaugural Board of the Ada Lovelace Institute, and became the UK Director of the new China-UK Research Centre for AI Ethics and Governance. Before moving to Cambridge in 2011, he was the ARC Federation Fellow and Challis Professor of Philosophy at the University of Sydney, where he was the founding Director of the Centre for Time. He is a Fellow of the British Academy, a Fellow and former Member of Council of the Australian Academy of the Humanities, and a Past President of the Australasian Association of Philosophy. He is also a member of the Global AI Ethics Consortium (GAIEC), which was initiated by the IEAI in 2020.