



Reflections on AI

Q&A with
Christoph Bartneck and Dwain Allan

“Robots are not just machines. They represent us without being us.”

The TUM IEAI had the pleasure of speaking with Christoph Bartneck and Dwain Allan. Dr. Bartneck is an associate professor at the University of Canterbury, focusing on the effect of anthropomorphism on human-computer interactions. Mr. Allan is a doctoral researcher at the Human Interface Technology Lab New Zealand (HIT Lab NZ) at the University of Canterbury.

Prelude

We cannot speak for the whole of AI. We are working in Human-Robot Interaction. This is an embodied form of artificial intelligence that is targeted at robots that interact socially with humans.

1. What is the biggest misconception about Artificial Intelligence?

Dwain: There are many misconceptions about embodied AI such as social robots. One misconception, with respect to robots in the workforce is that robots will only take over jobs where the work is either dull, dangerous, or dirty, leaving humans to focus on more creative work. The common thought is, that if you do work that is creative, such as graphic design or art you are relatively safe from AI-driven job displacement. However, we already have robots capable of performing such roles, take Ai-Da for example, which is a robot that makes drawings, paintings, and sculptures. Also, Hanson Robotics' Sophia has created NFT digital artworks. This is not

to say, that robots will be creative in the same sense that humans are creative, rather that it is conceivable that Robots will adequately fulfil some of these roles in society.

Christoph: Robots are not just machines. They represent us without being us. Our views on robots has been clouded by wishful thinking and science fiction. Pretty much like the Metaverse.

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

Dwain: While there is no consensus on what the most important question in social robotics and HRI related ethics is, if we take ethics in its broadest sense, to refer to principles discerning between behavior that helps vs. behavior that harms. Then one important, yet long-standing question is how do we ensure that robots are designed in such a way that they support, and enhance human social relationships rather than replace them, which as some have predicted, may lead to undermining the integrity of human relationships, and host of anti-social behaviours.

Christoph: Robots do not have to be like us to interact with us. Therefore it isn't necessary to build an electronic human brain or androids. Satisficing, is a portmanteau (portmanto) of satisfy and suffice which was introduced by Herbert Simon. We only have to satisfice, not come up with an eternal and absolute solutions for ethics in AI. It is not necessary to have a

mathematical proof to make it work. In particular since some of the ethical appraisal is in the eye of the beholder.

Ethics in AI is also only one problem of many. To achieve a certain functionality at all requires satisficing. Ethics is then only a variation of problem solving. It is not something inherently different. The real challenge is also not in the ethical rules themselves. If a robot would simply obey the law of the land, then the majority of rules are already defined. Although the frequently reference to the "reasonable woman/man" still introduces some ambiguity.

The real problem is in the necessary knowledge about the world. This knowledge is necessary for the robot to function at all. The robot not being able to function well will be the much more annoying aspect of robots compared to the few ethical decision it will have to do. Many of the robots functions will be mundane, with the notable exception of autonomous vehicles. Still, the AV has to be able to drive at all before we can consider more challenging ethical situations other than „do not crash“.

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

Dwain: This is largely a political question, as it is, in essence, about what sort of world we are bringing forth in developing social robots for everyday life, therefore, it should be resolved democratically. More practically, the development of ethical frameworks and standards would ideally be a participatory process involving a diverse group of specialists, such as, technologists, scientists, and philosophers,

as well as the intended community of users and those who may be directly affected.

Central to this process, however, is the role of the designer or those who can understand the multiple stakeholders, and tame the complexity of these frameworks shaping them with the appropriate intellectual clarity and effectiveness, so they may be utilized by regular human beings.

Christoph: The standard answer is „Everybody“. But this fails to acknowledge that experts have more and better insights into the challenges and solutions for ethical questions around AI. Still, the general public must be involved in the lawmaking similar to any other legal change.

The real questions is, why is currently not everybody involved? One challenge is that the topic at hand is inherently difficult and not everybody is even able to make meaningful contributions. We have to be grateful for those who at least are able to understand the complexities.

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

Dwain: The role of universities and other research institutions is to provide research expertise, clarify issues and help identify problems.

For this purpose, however, such institutions must be prepared to work more closely with industry than is perhaps currently the norm.

Christoph: I am not sure if the proximity to industry is a good thing. While knowledge transfer is important,

universities need to be able to maintain an objective point of view and money from companies tend to corrupt science. The government needs to strength the

universities to have academic freedom, including the means to finance it. Who else in society would be able to speak the truth without fears of losing their jobs?

5. How does the physical anthropomorphism of a robot impact people's attitudes towards it and why?

Dwain: Humans perceive computers and similar technology as social actors, as entities with personalities not merely as objects or things. With respect to robots, this social actor perception is amplified as a direct consequence of physical anthropomorphism, which we might define as the combination of physical embodiment, movement, autonomy, or the capability of sensing or responding to social cues in human environments.

There are numerous interesting examples of how this anthropomorphism impacts people's attitudes toward robots. For example, in the US military, there are stories of soldiers risking their lives to save robots, accounts of emotional distress over destroyed robots, robots receiving purple hearts, and funerals with gun salutes.

Returning to the theme of job replacement, though, research suggests that when robots are anthropomorphized as overly humanlike and/or highly intelligent or capable, they threaten human safety, resources, and jobs, as well as human uniqueness and identity. In such cases, we view these robots as members of a highly competent outgroup and incredibly threatening.

Christoph: Humans have become familiar with all sorts of animals, including other humans for thousands of years. When we see a robot move and behave at its own accord we cannot help but consider it to be somewhat alive. Being in the world rather than on a screen strengthens this mental processing. Rationally, we know that it is just a thing that moves. But our rational brain has never been our strong suit.

6. We often say that AI is changing and transforming our world. To what extent is AI, and more specifically robots, changing us and our relationships as humans?

Fundamentally, social robots hold the promise of improving the way humans relate to one another. For example, countless studies demonstrate that robots can improve communication among groups of humans and thus can help them perform better on tasks.

Take the Paro seal robot, which has been shown to inspire conversation among nursing home residents. Equally, social robots have been shown to facilitate greater communication between teachers, parents, and children. On the flip side, we are already witnessing children shouting rude commands at digital assistants, like Alexa or Siri, which has caused some to fear that this behavior will affect the way children behave toward people.

There is also research indicating that some individuals have married their virtual agents, due in part, to their lack of success in human-human relationships. So again, the concern is that interactions such as these will be amplified with the introduction of social robots and this may negatively change human relationships.

Either way, it stands to reason that as these machines enter our social environments, they will, in turn, change the way humans act and relate to one another socially.

Christoph: When sex with robots become more enjoyable than sex with humans then our society will be doomed. Already today people touch their phones more often than their partners. The problem is that we can program robots to behave in any way we want. No exceptions. I just hope that we will become bored by robots in the same way that we are starting to become bored with social media. It is just not the real thing.

Meet the Experts

[Dr. Christoph Bartneck](#) is an associate professor at the University of Canterbury. He has a background in Industrial Design and Human-Computer interaction, and his projects and studies have been published in leading newspapers, journals and conferences. More specifically, he focuses on the effect of anthropomorphism on human-robot interaction. As a secondary research interest he works on bibliometric analyses, agent based social simulations, and the critical review on scientific processes and policies. In the field of design, Dr. Bartneck investigates the history of product design, tessellations and photography.



[Mr. Dwain Allan](#) is a doctoral researcher at the Human Interface Technology Lab New Zealand (HIT Lab NZ) at the University of Canterbury. He has a background in design, marketing, innovation, and venture strategy, with a special interest in technology-based entrepreneurship. His research interests are in the cross-sections of social science, design, technology, and consumer behavior. His present work in the field of social robotics examines the role of implicit self-theories in determining how individuals perceive, evaluate, and respond to, social robots.

