Conditiohumana.io

"Fixing discrimination in algorithms mirrors in many ways the difficulties of eliminating discrimination in any social system"

Interview with Professor Nayef Al-Rodhan

Alexander Görlach: While the world believes artificial intelligence is the next step in our evolution you just recently posted a link to an article on Twitter that highlights the ability of brain-to-brain communication in humans. Should we focus instead on the incredible potential we humans have still to unwrap rather than on AI?

Professor Nayef Al-Rodhan: I don't believe that this is an "either-or" kind of choice. Both areas of research are poised to make strides in the coming decades, and each comes with its own set of promises (and ethical challenges). Brain-to-brain communication, for instance, to the extent that it will develop as a full-fledged technology, could bring significant changes to the economy and how we work (allowing, for instance, individuals to cooperate and work together, even without speaking the same language since what would happen would be that electrical activity in two or more brains becomes synchronized, and that does not require verbal communication).

This kind of networked communication would also enhance performance in humans, allowing them to become aware of problems and solutions they would have not spotted by themselves. The potential uses of brain-to-brain communications are vast, which is why this kind of technology has very much caught the attention of the DARPA in the US. At the same time, as promising or revolutionary as brain-to-brain communication may be, it could not make up for some of the potential benefits and uses of AI and AI-enabled technologies. For instance, AI could improve the prediction of disease outbreaks, as well as monitor in real-time the spread of viruses – such as, for instance, by tracking social media communications, and in warfare, AI is seen as important for aiding to process huge amounts of data (such as drone video footage).

Importantly, AI can also help tackle some of the challenges and risks emanating from climate change, such as by improving monitoring of deforestation, improving energy consumption and climate predictions. The benefits of AI could be enormous for our future and it can contribute to creating more sustainable livelihoods (and enhance human security) but it is important to recognize it cannot be a silver bullet for all of humanity's challenges and that there must be limitations to its deployment.

Alexander Görlach: There seems to be kind of an agreement that humans and machines will continue to grow closer together. What would the next steps on this way entail according to you?

Professor Nayef Al-Rodhan: I believe that, in addition to the development of AI and smart robots – entities that are outside our bodies – another important trend will be towards more and more integration of technology *within* the body, as well as the use of neuro-technological and neuropharmacological means of enhancing our bodies. Some of the most radical and profound changes for the human condition will take place through such interventions. That is not to say that we

will not become more accustomed to or reliant on robots, but I want to draw the attention to human enhancement because its implications and risks in the long run are truly existential.

Moreover, as I have stated on numerous occasions, the roots of the challenges with human enhancement lie in our very nature, which is why it is going to be difficult to limit the appeal of enhancement technologies, even when we may 'rationally' recognize that such technologies will be harmful to us. Our neurochemical makeup predisposes us in some basic ways. There are five basic human motivators that drive human action and choices throughout our existence, which I previously called the *Neuro P5*: power, profit, pleasure, pride and permanency. If a technology appears which promises to enhances one, or better still, several of these human motivators, we will go in the direction of adopting that technology even if the implications are not in our favor in the long run. Or, enhancement technologies are predominantly premised on augmenting us and making us in some way or another stronger and smarter.

By their very nature, these technologies will appeal to the basic human motivators encoded in our neurochemical makeup – hence the importance of regulations to ensure that we stave off the most dangerous, irreversible and extreme forms of enhancement.

The benefits of AI could be enormous for our future and it can contribute to creating more sustainable livelihoods (and enhance human security) but it is important to recognize it cannot be a silver bullet for all of humanity's challenges and that there must be limitations to its deployment.

Alexander Görlach: When we speak about artificial intelligence we seem to be neglecting other key components that make us human, empathy for example. Empathy amongst others is a feature that enables us to act as a moral species, to make ethical judgements. Is there any way in which we could make our emerging technology a more moral one?

Professor Nayef Al-Rodhan: Any meaningful discussion of moral technologies and moral robots must acknowledge that we are very far from the moment when the question of teaching or embedding moral competence into machines becomes real. Because AI and computing technologies have indeed made outstanding progress in the past decades, there has been an exaggerated and sudden hype around the potential of life-like machines accompanying us in our daily lives as our co-workers, professors, or friends.

That said, the prospect of moral robots has already elicited some interesting views, and an interesting dichotomy has been proposed, separating "top-down" morality from "bottom-up" morality – the former refers to the set of values encoded in the machines by programmers, while the latter is an approach that would enable the robots to learn moral values by themselves and as they continue to exist, learn and evolve in (our) social reality. (I have discussed this in more detail here.) The crucial challenge with the 'bottom-up' approach consists in developing the kind of technology that would allow robots to learn moral competencies by themselves. Advances in neuromorphic technologies may open the path for such advances one day.

That still leaves some open-ended questions about how exactly that moral compass would consolidate and the limitations of 'robot morality'. The question of empathy, raised in your question, highlights for instance the limits of developing such emotions in technologies. Robots may acquire notions of right and wrong, and learn to operationalize them in complex situations, but will lack the affective brain

circuits that render many human emotions truly complex and nuanced. Empathy is linked to the so-called pain matrix, which refers to the brain areas that are involved in processing pain, such as the bilateral anterior insula, the dorsal anterior cingulate cortex, brain stem, and the cerebellum; empathy has also been shown, in more recent studies, to be linked to somatosensory processing – for instance, when we see the pain experienced by others.

While we can imagine technologies and AI to be able to acquire, one day, some moral competencies and even apply them in complex situations, the intricate mechanisms involved in moral judgment in the human brain renders human morality uniquely sophisticated, and hard to replicate in non-human mechanisms.

Alexander Görlach: One element of this discussion is how to immunize algorithms from our biases. Some argue that is utterly impossible as the sheer metric by which you look upon data is already biased as it depends heavily on the logic and linguistic framework of the computer scientist applying an algorithm to a specific problem set. Do you share this view? Why or why not?

Professor Nayef Al-Rodhan: The question of biases in algorithms is a very serious one and I believe it is going to be very difficult to 'immunize algorithms' from biases. We see with more frequency reports of cases where algorithms are imbued with bias that sways decisions in the justice system, health care, and other fields. Just last October, another case of racial bias in algorithms was revealed in the US, where the software used across hospitals in the country relied on an algorithm that was less likely to refer black people to personalized care, as opposed to white people. However, correcting those algorithms, and clearing them of bias, is not an easy and clear-cut process either.

In fact, fixing discrimination in algorithms mirrors in many ways the difficulties of eliminating discrimination in any social system. One much-discussed solution would be to significantly increase diversity among programmers and algorithm designers but even so, it is misplaced to expect a group of individuals, no matter how racially or ethnically diverse, to predict or mitigate all the possible risks in automation. There are other challenges to address too, such as how the data is collected and the quality of that data (Is it already imbued with prejudice? Does it reflect the vision of a tolerant and inclusive society? etc.), the purpose underlying the algorithm (defining exactly what the deep-learning model is supposed to do), or finally the varied social norms and different responses across communities – for instance, perception of what is fair or acceptable sometimes differs across communities.

Alexander Görlach: I feel talking about artificial intelligence with moral agency and robots with rights comparable to human rights will not have much use if we do not also at the same time advance in our understanding of what the human being is. In your opinion what should anthropology of our time entail given the advancements of the last decades that clearly render large parts of traditional anthropology obsolete?

The question of biases in algorithms is a very serious one and I believe it is going to be very difficult to 'immunize algorithms' from biases. We see with more frequency reports of cases where algorithms are imbued with bias that sways decisions in the justice system, health care, and other fields.

Professor Nayef Al-Rodhan: The "human" element will be doubly transformed by the advent of new technologies. On the one hand, the increasing sophistication of AI, and one day, the existence of robots

with advanced moral competencies, will transform our social relations and also how we regard ourselves as humans; on the other hand, technologies of human enhancement will further transform us. While enhancement may boost cognitive and physical capabilities, they may in the process diminish some deeply human features that have been pivotal to us as a species both for survival but also for cooperation and prosperity.

In other words, these technological advances will empower us in some ways, while simultaneously disempower us in other profound ways. "Anthropology of our time" will need to account for the ways in which technology is transforming notions of the 'self' and of citizenship and belonging – especially as AI may enable more surveillance and therefore a different kind of relationship to power.

Moreover, advancements in recent decades also point to important transformations to human institutions and values, such as meritocracy (enhancements may confer extra capabilities to those who can afford access to such interventions), the authenticity of the human experience, as well as the notion of equal human rights and human dignity grounded in our shared humanity. It is perhaps more accurate to say that the study of anthropology for our times will be rebranded into 'techno-anthropology' because technology and innovation will feature heavily in all our social relations, self-identity and even our sense of belonging to groups and states.

Alexander Görlach: On the grounds of such a new image of mankind: can technology help us address and solve the most pressing issues of our time: climate change and a new economic system that ultimately creates fairness in an age of shrinking resources? This development would effectively bring the third challenge of the current era to a halt to: the millions of refugees that flee due to climate change and desolate economies in their home countries.

Professor Nayef Al-Rodhan: Yes, technology can help us tackle some of the most pressing challenges of our times, including in the area of climate change. Some suggested possible ways to use machine learning to that end include, among others, improving energy efficiency and making buildings more energy-efficient, monitoring deforestation, and creating better climate predictions. However, the potential of technology to be a savior of humanity is ultimately limited and must not be used with a view to compensate for lack of human responsibility and responsible human action.

NAYEF AL-RODHAN AL-RODHAN

Prof. Nayef Al-Rodhan (@SustainHistory) is a Neuroscientist, Philosopher and Geostrategist. He is an Honorary Fellow at St Antony's College, University of Oxford, and Senior Fellow and Head of the Geopolitics and Global Futures Programme at the Geneva Centre for Security Policy, Geneva, Switzerland. Through many innovative books and articles, he has made significant conceptual contributions to the application of the field of neurophilosophy to human nature, history, contemporary geopolitics, international relations, cultural studies, future studies, and war and peace.