Science SEPTEMBER 4, 2019 11:27 PM AEST

Share Twitter Facebook

Artificial intelligence: only way is ethics

During his talk, Nallur called for increased collaboration between computer scientists, legal professionals and experts in the domains where AI technologies are being applied

CERN has an ambitious upgrade programme for its *flagship accelerator complex* over the next two decades. This is vital to continue pushing back the frontiers of knowledge in fundamental physics, but it also poses *some gargantuan computing challenges*.

One of the potential ways to address some of these challenges is to make use of artificial intelligence (AI) technologies. Such technologies could, for example, play a role in filtering through hundreds of millions of particle collision events each second to select interesting ones for further study. Or they could be used to help spot patterns in monitoring data from industrial control systems and prevent faults before they even arise. Already today, machine-learning approaches are being applied to these areas.

It was in view of the potential for further important developments in this area that *Vivek Nallur* was invited to give a talk last week at CERN entitled '*Intelligence and Ethics in Machines – Utopia or Dystopia?*'.

Nallur is an assistant professor at *the School of Computer Science at University College Dublin in Ireland*. He gave an overview of how AI technologies are being used in wider society today and highlighted many of the limitations of current systems. In particular, Nallur discussed challenges related to the verification and validation of decisions made, the problems surrounding implicit bias, and the difficulties of actually encoding ethical principles.

During his talk, Nallur provided an overview of the main efforts undertaken to date to create AI systems with a universal sense of ethics. In particular, he discussed systems based on consequentialist ethics, virtue ethics and deontological ethics – highlighting how these can throw up wildly different behaviours. Therefore, instead of aiming for universal ethics, Nallur champions an approach based on domain-specific ethics, with the goal of achieving an AI system that can act ethically in a specific field. He believes the best way to achieve this is by using games to represent certain multi-agent situations, thus allowing ethics to emerge through agreement based on socio-evolutionary mechanisms – as in human societies. Essentially, he wants AI agents to play games together again and again until they can agree on what actions should or shouldn't be taken in given circumstances.

"We shouldn't try to jump from no ethics in AI to universal ethics; let's take it step by step," says Nallur. "To start, we should aim to have systems that work and can have liberty within specific domains. To achieve this, we will need intense and fundamental collaboration between computer scientists, domain experts and legal professionals."

Nallur was invited to speak at CERN by *CERN openlab*, which is running *a number of R&D projects related to AI technologies* with its industry and research collaborators. "Naturally, CERN doesn't have to deal with the kind of ethical quandaries that those using AI in a medical or law-enforcement context face," says Alberto Di Meglio, head of CERN openlab. "However, it would be a mistake to dismiss this as simply an interesting philosophical exercise in the context of particle physics. Here at CERN, we are proud that tools and techniques we develop are often adopted for use by other communities – within both research and industry. As such, it is important to think about ethical considerations related to AI technologies from a very early stage." He continues: "I hope that this fascinating talk will serve to ignite further discussion within our community."

(Image: Andrew Purcell/CERN)