UKRI Centre for Doctoral Training in Safe & Trusted AI

Overview

The UKRI Centre for Doctoral Training (CDT) in Safe and Trusted Artificial Intelligence (STAI) brings together world leading experts from King's College London and Imperial College London to train a new generation of researchers in safe and trusted artificial intelligence (AI).

The STAI CDT offers a unique four-year PhD programme, focussed on the use of model-based AI techniques for ensuring the safety and trustworthiness of AI systems. Students will engage in various training activities, alongside their individual PhD project, ensuring that not only are they trained in state-of-the-art AI techniques, but also that they acquire a deep understanding of ethical, societal, and legal implications of AI in a research and industrial setting. Through engagement with the CDT's diverse range of industrial partners, students will be exposed to the different experiences, challenges, and technical problems involved in both startups and large corporations.

The CDT leadership team is committed to providing an inclusive environment in which diverse students can thrive. Diversity is crucial for enabling world leading research, impact and teaching, and an inclusive environment is vital to allow people to contribute their best.

What is Safe & Trusted AI?

AI technologies are increasingly ubiquitous in modern society, with the potential to fundamentally change all aspects of our lives. While there is great interest in deploying AI in existing and new applications, serious concerns remain about the safety and trustworthiness of current AI technologies. These concerns are well-founded: there is now ample evidence in several application domains (autonomous vehicles, image recognition, etc.) that AI systems may currently be unsafe because of the lack of assurance over their behaviour. Even in areas where AI methods function to high standards of correctness, there remain challenges. AI decisions are often not explained to users, do not always appear to adhere to social norms and conventions, can be distorted by bias in their data or algorithms and, at times, cannot even be understood by their engineers. An AI system is considered to be safe when we can provide some assurance about the correctness of its behaviour, and it is considered to be trusted if the average user can have confidence in the system and its decision making.

To read more and apply, visit: https://www.kcl.ac.uk/study/postgraduate/research-courses/safe-and-trusted-ai-phd