AI + X Summit 2024

Track «Responsible AI» – October 4, 2024

Notes on the event

At the AI+X Summit on October 4, 2024, experts from the University of Zurich (UZH) and the Kanton of Zurich met to discuss the problem of «Responsible AI». In the track/workshop of the same name, Markus Christen, Florent Thouvenin and Nicolo Pagan from the UZH presented philosophical, legal and informatics perspectives on responsible AI in order to answer questions about who (if anyone) can be held responsible for damage caused in particular human-AI interactions; what psychological dispositions humans have that enable them to be held responsible in these cases; and how to deal with AI systems legally, socially or politically, taking into account questions of responsibility.

Data was presented showing that humans show a disposition to attribute moral responsibility to AI. While noting that responsibility is not a legal term, it was pointed out in legalese that AI cannot be held responsible. Different strategies to regulate AI were explored including the EU's AI Act. This takes a more risk-based approach which depends on defining what AI is in advance, as a result of which it would need to be updated as AI evolves and changes. It was proposed that a technology-neutral approach, that focuses on harms rather than risk reduction, would be preferable as a basis for a Swiss law.

The other side of the coin to legal regulation is technical solutions. In this context principles including fairness of algorithms, transparency of systems, privacy, safety and robustness, and accountability were framed. AI systems, for example, should not be trained on biased data, and their decision-making process should be traceable and justifiable. Building responsible systems requires collaborative efforts across industry, academia, and policy.

Following group discussions on the above issues, two practice input sessions were held which saw, first, Raphael von Thiessen from the Office for Economy, Programme Manager AI Sandbox, at the Canton of Zurich, presenting on the Canton's «Innovation Sandbox for Artificial Intelligence». The sandbox brings together various stakeholders from public administration, private sector, and the research community and provides an innovative environment for testing and developing AI technologies and ensuring responsible AI. Five projects have been successfully completed in the area of smart parking, autonomous systems, automated infrastructure maintenance, machine translation and AI in education, and insights shared with public administration, business, and research. Twenty-four projects have been submitted from sectors including construction, health, and education.

Secondly, Elena Maran, Global Head FS & RAI, and Pierre Oberholzer, Lead Data Scientist, ended the «Responsible AI» discussions by presenting their Modulos platform. Pointing out the importance of AI governance in the face of exploitation and other bad practices that it can engender, they also note that complying with the ethical guidelines of AI can be an overwhelming process, due to multiple regulatory frameworks, such as Singapore's MAS FEAT and the EU AI Act. Modulos aims to guide organizations through the process, enabling them to manage risk and ensure compliance, all with a view to ensuring that responsible AI practices are implemented.