Interaction-Centred Design for Responsible Human-AI Symbiotic Partnership: The Next Stage of Evolution

With collective human-machine intelligence, the human-machine symbiosis (HMS) technologies are prevalent in society today and capable of solving complex problems. However, the trend raises important questions about the complications, liabilities, risks, and trust associated with increasing intelligence and adaptivity in these human-machine systems given both humans and machines have limitations. It is even more challenging when we are facing insufficient data, indeterministic conditions, and inexhaustive solutions for uncertain actions.

Recent accidents to the Boing 737 Max passengers ring the alarm again about the importance of the appropriate design methodologies and regulations for these safety-critical socio-technical systems. It is not only about the design of the system, the procedures, the processes, the training, and certification, etc., but a systematic approach when we design, develop, verify, validate, and regulate those systems, including but certainly not limited to aviation, process control, systems with certain levels of automation as well as artificial intelligence (AI). It is not only about the safety of those systems, but more importantly human lives. Today, Human-Centred Design (HCD) approach is obviously not sufficient to address the interaction issues when transferring the control authority from human to increasingly powerful AI-enabled decision-making functionalities. A new design paradigm is imperatively needed for the safe and responsible human-AI symbiotic partnership.

This lecture aims to provide a state-of-the-art of design strategies to address broader issues when humans transition their interactions with AI/Autonomy from “on-the-loop” to “in-the-loop”. A context-based and interaction-centred design (ICD) approach for developing a safe and responsible partnership between human and AI to support the broader applications of advanced AI-enabled socio-technical systems will be introduced. The ICD framework, its analytical techniques, design methodologies, implementation strategies, and test and evaluation processes have helped the scientific and defence communities understand the optimal means by which human operators can be teamed up with autonomy and AI to conduct missions successfully in complex environments. It became a guiding principle and strategy for the development of international standards and a United Nations White Paper to address human-automation interaction issues.

This lecture will explain the ICD approach in detail as the guidance on solutions to address a variety of design, development, and operational issues of human-AI symbiosis technologies. Validation studies on ICD utility and effectiveness through real-world technology evaluation activities will be reviewed. A related trust model IMPACTS (Intention, Measurability, Performance, Adaptivity, Communication, Transparency, and Security) will also be introduced as an enduring strategy and appropriate solution to address human-AI trustworthy issues. Challenges in integrating the ICD approach into systems engineering and validation processes will then be discussed for the future directions in R&D and exploitation of human-AI symbiosis technologies.