Human-Centered AI to foster Trustworthy AI

Andreas Holzinger, Human-Centered AI Lab,

University of Natural Resources and Life Sciences Vienna, Austria

We are again in an AI spring, perhaps even already in an AI summer, thanks to the successes of statistical probabilistic machine learning. However, if we look at the world of robotics or cyber-physical systems as an example, we will find that often seemingly simple problems are not solvable or only insufficiently solvable. This is directly related to robustness, because perturbations in the input data can have dramatic effects on the output and lead to completely different results. This is relevant in all critical domains where we work with real data from our environment, i.e. where we do not have i.i.d. laboratory data. Therefore, the use of AI in real domains that impact human life (agriculture, climate, forestry, health, etc.) has led to an increased demand for trustworthy AI. In sensitive domains where traceability, transparency, and interpretability are required, explainable AI (XAI) is now even essential due to regulatory requirements. One approach to making AI more robust is to combine statistical learning with knowledge representations. And this is where interactive machine learning can help. For certain tasks, it can be beneficial to include a human in the loop. A human expert can sometimes - not always, of course - bring experience and conceptual understanding to the AI pipeline. Such approaches are not only a solution from a legal perspective, but in many application areas the question of "why" is often more important than a pure classification result. Consequently, both explainability and robustness can promote reliability and trust and ensure that humans remain in control, thus complementing - rather than replacing - human intelligence with artificial intelligence.

Andreas Holzinger pioneered in interactive machine learning with the human-in-the-loop promoting robustness and explainability to foster trustworthy AI. He advocates a synergistic approach to Human-Centered AI (HCAI) to align artificial intelligence with human values, ethical principles, and legal requirements to ensure privacy, security, and safety. For his achievements he was elected a member of Academia Europaea in 2019, the European Academy of Science, of the European Laboratory for Learning and Intelligent Systems (ELLIS) in 2020, and Fellow of the international federation of information processing (ifip) in 2021. He obtained his Ph.D. with a topic in Cognitive Science from Graz University in 1998, and his Habilitation (second doctorate) in Computer Science from Graz University of Technology in 2003. Andreas was Visiting Professor for Machine Learning & Knowledge Extraction in Verona (Italy), RWTH Aachen (Germany), and the University College London (UK). From July 2019 until February 2022 Andreas was a Visiting Professor for explainable AI at the University of Alberta (Canada). Andreas Holzinger has been appointed full professor for digital transformation in smart farm and forest operations at the University of Natural Resources and Life Sciences Vienna as of March 1, 2022 and is head of the Human-Centered AI Lab Vienna.

https://www.aholzinger.at/