

From Explainable AI to Human-Centered AI

Abstract: In the medical domain the expectations to automatic AI systems are high, particularly in disciplines requiring prognostic models (oncology) and/or decision support (radiology, pathology). Due to the raising ethical, social, and legal issues governed by the European Union, the field of explainable AI is becoming extremely important. The problem of explainability is as old as AI itself, and classic rule-based approaches have been comprehensible "glass-box" approaches. Nevertheless, their weakness was in dealing with non-linearities and the intrinsic uncertainties of medical data. The progress of probabilistic machine learning, the availability of big data and computational power has made AI successful today, and in certain medical tasks deep learning even exceed human performance. However, such approaches are considered as "black box"- models, and even if we understand the underlying mathematical principles of such models, they still lack explicit declarative knowledge. Consequently, in the future we need context-adaptive procedures, i.e. systems that construct contextual explanatory models for classes of real-world phenomena. One possible step is in linking probabilistic learning methods with large knowledge representations (ontologies), thus allowing to understand how a machine decision has been reached. Our aim is to make machine decisions re-traceable, interpretable and comprehensible. The aim is to explain why a certain machine decision has been reached because the "why" is often more important than the mere classification result. The re-traceability and interpretability on demand shall foster reliability and trust ensuring that the human remains in control, so to augment human intelligence with artificial intelligence and vice versa.

Bio: Andreas Holzinger is lead of the Human-Centered AI Lab (Holzinger Group) at the Medical University Graz and since 2016 he is Visiting Professor for machine learning in health informatics at Vienna University of Technology. Andreas was Visiting Professor for Machine Learning & Knowledge Extraction in Verona; the RWTH Aachen, University College London and Middlesex University London. He serves as consultant for the Canadian, US, UK, Swiss, French, Italian and Dutch governments, for the German Excellence Initiative, and as national expert in the European Commission. He is in the advisory board of the Artificial Intelligence Strategy "AI Made in Germany 2030" of the German Federal Government and in the advisory board of the "Artificial Intelligence Mission Austria 2030". Andreas Holzinger promotes a synergistic approach to Human-Centred Artificial Intelligence (HC-AI) and has pioneered in interactive machine learning (iML) with the human-in-the-loop. Andreas' goal is to augment human intelligence with artificial intelligence to help to solve problems in health informatics. Andreas obtained a Ph.D. in Cognitive Science from Graz University in 1998 and his second Ph.D. in Computer Science from TU Graz in 2003. He serves as Austrian Representative for AI in IFIP TC 12, he is organizer of the IFIP Cross-Domain Conference "Machine Learning & Knowledge Extraction (CD-MAKE)" and is member of IFIP WG 12.9 Computational Intelligence, the ACM, IEEE, GI, the Austrian Computer Science and the Association for the Advancement of Artificial Intelligence (AAAI). More information: <https://www.aholzinger.at>