ICRES 2019: International Conference on Robot
Ethics and Standards, London, UK, 29-30 July 2019.
https://doi.org/10.13180/icres.2019.29-30.07.p04

WHAT DOES IT MEAN TO TRUST AI SYSTEMS?

RAJA CHATILA
Sorbonne University, France

The High-Level Expert Group on Artificial Intelligence (AI HLEG) appointed by the EU Commission is tasked with elaborating recommendations on ethical, legal and societal issues related to AI and on policy and investments. The group has produced a document released in April 2019 entitled "Ethics Guidelines for Trustworthy Artificial Intelligence".

Computerized socio-technical systems, based on algorithmic computations and decisions that impact human individuals and society in a way or another, must indeed be trustworthy, and AI systems are no exception. The academic and industrial communities developing software-based systems have produced several techniques to achieve their "dependability" or "resilience". Software validation and verification techniques, such as error detection and recovery mechanisms, model checking, detection of incorrect or incomplete system knowledge, and resilience to unexpected changes due to environment or system dynamics have been developed for several critical applications such as aviation.

However, as decisions usually devoted to humans are being more and more delegated to machines, sometimes running computational algorithms based on learning techniques, ethical considerations become central. New issues have to be considered technically, in the design process and in the governance of these systems, and properties such as transparency, accountability and explainability take more importance.

