

# Human-aware AI

*Tuesday, 13 October 2020 09:00 (1 hour)*

Machine learning is often synonymous with predictive models of exceptional accuracy. In classification they are commonly evaluated with summary measures of predictive performance - but is this enough to validate a complex algorithm? Non-linear models will exploit any artefacts in the data, which can result in high performing models that are completely spurious. Examples of this will be shown. This leads onto the need for a clear ontology of model interpretability, for model design and usability testing. This reinforces the emerging paradigm of AI not as a stand-alone oracle but as an interactive tool to generate insights by querying the data, sometimes called xAI or AI2.0 –AI with a person in the loop.

In this talk, Professor Paulo Lisboa will describe how probabilistic machine learning models can be presented as similarity networks and how SVMs and neural networks generate simpler and transparent models including globally accurate representations with nomograms. Perhaps surprisingly, this can buck the accuracy/interpretability trade-off, by producing self-explaining neural networks that outperform black box models and match deep learning. The dependence on the main predictive variables will be made explicit for a range of benchmark data sets commonly used in the machine learning literature.

Paulo Lisboa is Professor in the Applied Mathematics at Liverpool John Moores University, UK and Project Director for LCR Activate, an ERDF funded £5m project to accelerate the development of SMEs in the Digital Creative Sectors in the Liverpool City Region. His research focus is advanced data analysis for decision support, in particular with applications to personalised medicine and public health. His research group on data science has developed rigorous methodologies to make machine learning models interpretable by end users.

**Presenter:** LISBOA, Paulo