## Data - Knowledge Environment and Knowledge Landmarks

## in Machine Learning

Witold Pedrycz Department of Electrical & Computer Engineering University of Alberta, Edmonton, Canada wpedrycz@ualberta.ca

## Abstract

The unpreceded progress in Machine Learning (ML) can be attributed to an efficient use of masses of data as being recently exemplified through numerous constructs of LLMs and foundation models.

It becomes intriguing, though, that while exhibiting a heavy reliance on data, a role of knowledge in ML has not been clearly considered. In this talk, we advocate an ultimate importance of synthesizing a unified design knowledge-data (KD) of Machine Learning or KD-ML, for brief. As a new paradigm, KD-ML focuses on a prudent and orchestrated engagement of data and knowledge in the design practices in the area.

The fundamentals of the KD environment are formulated along with a historical perspective and the key highlights are identified. The issues of origin of problem-oriented knowledge, taxonomy of knowledge and the and its main features are discussed.

Data and knowledge arise at very different levels of abstraction with knowledge being formalized and represented at symbolic level. This constitutes a genuine challenge as data are predominantly numeric. We stress that in the development of a cohesive and unified framework of coping with data and knowledge in learning processes, one needs to reconcile highly distinct levels of abstraction (numeric-qualitative) and with this regard information granules play a pivotal role.

We offer a taxonomy of knowledge by distinguishing between scientific and common-sense knowledge and elaborate on a spectrum of ensuing knowledge representation scheme. In the sequel, the main categories of knowledge-oriented ML design are discussed including physics-informed ML (with the reliance of scientific knowledge), an augmentation of data driven models through knowledge-oriented constraints (regularization), a development of granular expansion of the data-driven model and ways of building ML models in the presence of knowledge conveyed by rules. When analyzing the proposed categories, it is also clearly explained how the new ML environment helps avoid a detrimental effect of *data blinding*. Selected schemes of the KD unified environment and ensuing learning schemes are discussed including a study on LLM-based knowledge acquisition.