The Past, Present, and Future of Artificial Intelligence:
From Black-box to White-box Deep Networks

Yi MA

Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, US;

Data Science Institute and Department of Computer Science, The University of Hong Kong, Hong Kong

Email: mayi@hku.hk

In this talk, the speaker will provide a more systematic and principled view about the practice of artificial intelligence in the past decade from the history of the study of intelligence. He argues that the most fundamental objective of intelligence is to learn a parsimonious and structured representation of the sensed world that maximizes the internal *information gain*. This objective naturally leads to a unifying computational framework which integrates fundamental ideas from information theory, optimization, feedback control, and game theory, hence connects us back to the true origin of the study of intelligence 80 years ago. We contend that this new framework provides a unifying understanding and purely white-box explanation for almost all recent and current practices of artificial intelligence based on deep networks, including CNN, ResNet, and Transformers. Probably most importantly, it reveals out a much broader and brighter future for developing next-generation autonomous intelligent systems that truly emulate the computational mechanisms of natural intelligence.