

「データ駆動社会における AI の信頼性と意思決定の未来」

竹内 孝 (京都大学大学院情報学研究科)

アブストラクト：

デジタル化が進む現代社会では、人工知能（AI: Artificial Intelligence）が重要な役割を果たすことが期待されています。近年、AI はビッグデータからパターンを学習し、未知のデータから予測する機械学習や深層学習によって飛躍的に進化しています。これにより、AI の予測を参考にした合理的な意思決定や重大なインシデントの予防が可能になりつつあります

しかし、現在の AI は予測精度に限界があり、そのシステムや結果の信頼相当性(Trustworthiness)には課題が残ります。たとえば、チャット形式で会話する大規模言語モデル（LLM: Large Language Model）は、事実に基づかない情報を生成する「幻覚（ハルシネーション）」を起こすことがあります。

それでも、AI を活用しないという選択肢は現実的ではありません。

本講演では、人間と AI の間に信頼(Trust)を構築し、AI の信頼相当性を向上させる方法について探ります。そして、未来の意思決定を AI により支援する社会の展望を紹介します。

以上