

科学技術の潮流

15

JST研究開発戦略センター

2つの潮流

人工知能（AI）技術は、さまざまな応用が社会に広がるにつれて、精度・性能の向上だけでなく、安全性・信頼性の確保を強く求められるようになってきた。例えばAIのブラックボックス性は、判定結果の理由が説明されず、事故が起きたときに責任の所在が明らかにならないとか、差別・不公平を生むような学習がされていないかといった懸念を生んでいる。また、AIの誤認識を誘発させる攻撃が可

能であるとか、人間には見分けるのが困難なフェイク画像・動画・文章を簡単に作れるようになってしまったという悪用の懸念も強まっている。このような「新世代AI」の懸念に対処し、「信頼されるAI」を目指すのがAI研究の第1の目標である。このように基礎的な研究も立ち上がってき

た。これが第2の潮流である。この取り組みは個別の問題への対処を積み上げる傾向にあるのに対して、現在のAIの中心である深層学習に、学習したことを組み合わせることで別な状況にも応用できる。このような人間の知能に関する基礎研究の成果・知見がヒントになり得る。

「信頼されるAI」にむけては、AIシステム「新世代AI」への

信頼される新世代AI研究

日本の勝ち筋

「信頼されるAI」になり得るはずである。日本は、AI開発において日本の強みに



科学技術振興機構（JST）研究開発戦略センターフェロー（システム・情報科学技術ユニット） 福島 俊一

東京大学理学部物理学科卒、NECCにて自然言語処理・情報検索の研究開発に従事後、2016年から現職。工学博士。11・13年東大大学院情報理工学系研究科客員教授、21年情報処理工学フェロー。

AI研究2つの新潮流		
	信頼されるAI	新世代AI
AI研究の潮流	AIシステムの安全性・信頼性を確保するための研究	現在の深層学習の限界を克服する新しいAIの仕組みの研究
日本にとってのチャンス	従来から品質は日本の強み、AI分野でも具体的な取り組みで先行	まだ初期ステージなので日本も先行チャンス、ロボットや知能基礎研究の強み活用

JST研究開発戦略センター「人工知能研究の新潮流～日本の勝ち筋～」(2021年6月)を基に作成

取り組みは米国国防高等研究計画局（DARPA）が「AIネクスト」を掲げた研究投資を推進しているが、融合や脳科学・認知発達などの知能基礎研究の強みの活用も見込める。深層学習を中心とした現在のAIは、米中2強と言われる状況だが、ここで述べた2つの新潮流で先行することで、日本の存在感が高まると期待したい。（金曜日掲載）