

書評

佐藤裕著

『人工知能の社会学——AIの時代における人間らしさを考える』

(ハーベスト社、2019年)

本田 康二郎

2000年代から始まった第3次AIブームは、自然言語処理、画像認識、学習などのこれまでになかった新しい機能を有する人工知能をうみだした。この新しいタイプの人工知能は、研究室の中だけにとどまらず、商品化されて一般社会にも普及しはじめている。その影響は、検索エンジン、ゲーム産業、医療診断、セキュリティ、自動車産業、金融工学、ロボット工学、芸術など様々な方面に及んでいる。コミュニケーション能力を有するロボットが老人の心理的ケアに活用されるようになり、人工知能が人間関係のネットワークの中で役割を持つようになっている。さらに、自動運転車が普及すれば、我々の日々の行動を人工知能が媒介するようになる。この新技術によって、我々の社会生活はどのように変容していくのだろうか。

このような考察を進めるためには、人間行動の社会性を考察してきた社会学の知見が必要になってくるであろう。『人工知能の社会学』は、社会学者が試みる人工知能批判として、国内では嚆矢となる作品である。佐藤は、人間の行動を「志向性」と「言語ゲーム」の二つの概念を用いて特徴づけている。「志向性」とは「方向づけ」を意味し、これは「活性化」、「選択肢」、「評価基準」という三つの条件で置き換えることができるという (p. 29)。人間の行動は、何等かの目的に向かって活性化されており、またその目的に至るための選択肢が含まれており、さらに選んだ選択肢が適切であったかを測る評価基準を備えている。これらの三つの条件を備えた行為主体は「志向的」であると表現される。そして、佐藤の定義では「言語ゲーム」とは「言語によって志向性を与えられたいと欲する」(p. 39) とされる。言語ゲームには大きく分けて、「質問・応答の言語ゲーム」と「命令・行為の言語ゲーム」の2種類があるという。人間

の行っているそれぞれの言語ゲームの特徴を分析し、その特徴とAIが行う推論や判断を比較検討していくことが本書の主な内容になる (第2部および第3部)。

目の前にあるものをリングだと見分け、それを食べられると判断することは人間にはたやすい。例えばこの程度の信頼性があればYESと判断できるような常識 (社会的ルール) を人間は持っている。しかし、AIにこれをやらせるためには判断のための厳密なルールを外部から設定しなければならない。どんな場合でも適用できる厳密なルールがなかった場合は、AIには判断が下せないということになる。ところで、ルールがなかったとしても、人間ならば「責任をとる」という意思決定を行うことができる。例えば、食べられると判断して目の前のものを食べ、味が悪かったり、おなかを壊したりしても、人間ならばそれは仕方ない (自己責任だ) と考えることができる。しかし、AIは責任をとるということができないので、曖昧な情報をもとに行動を起こすことができない。また、人間は誰か信頼できる他者の意見を信じて行動を起こすことができる。それはリングだし食べることができるよ、という信頼できる人からの意見を信じれば、人間は目の前のものを食べることができる。ところが、AIはそもそも誰が信頼できるかを判断できない。佐藤によれば、人間の何気ない日常の判断や行動は、常識とか信頼といった社会的な構成物によって媒介されているのが常である。裏を返せば、人間と同じことを行うことは、社会性を持たない計算機 (= 人工知能) にとっては難しいことなのである。

人間はある目的を達成するために選択肢を吟味し、行動を選び、時には失敗をする。そして、その失敗から「学習」をすることができる。ところが、AIの失敗は確率的な失敗であり、そもそも何らかの目的を志向する中で過ちを犯したわけではない。つまりAIには人間的な学びの機会がないわけである。そして、AIが高度化すればするほど、その計算過程が不可視となっていく。AIの失敗は人間にとって理解しがたいものとなっていく。現状のAIのままで、例えば自動運転車を実装したとすれば当然混乱が生じてくるであろう

う。AIが人間社会の一員として受け入れられるためには、それは人間同様の「利害」や「意志」を持った志向的に行動できる存在とならなければならない（第10章）。言いかえると、現状のAIに対する人間の利点がどこにあるかといえば、過ちを犯す能力を持つことだということになる。人間は自らの責任でリスクを冒して挑戦することができるし、あるルールが現状に適合しなくなればそのルールを破ることもできるわけである（第12章）。

本書は、人間行動の社会性という観点から、現状のAIの弱点を指摘するとともに、機械との違いを強調する中で、人間の特徴や美点を浮き彫りにすることに成功している。ところが、その説明原理として活用している「志向性」や「言語ゲーム」といった現代哲学の主要なキーワードが、佐藤独自の定義のもとで使われており、それが本書の読解を難しくしている。現象学や分析哲学の先入観をもって読もうとすると、本書の趣旨はなかなかつかめなくなるので注意が必要である。これらの概念が佐藤独自の術語で再定義されれば、本書で表現された粘り強い論考がさらに分かりやすくなるのではないだろうか。