



《第54回》 ロボット・AIは「他者」たりうるか

松井 哲也

1. 筆者の研究背景

筆者の研究上の関心は、AI研究の分野における他者モデル・XAI（説明可能AI）への疑義、およびそれら乗り越えた人—“他者”インタラクション研究の構築である。

そもそも筆者は、博士後期課程までは理学分野で生物行動学の研究を行っていた。そのときに、動物の群れ行動を対象として研究を行う上で抱いた直感が、現在まで筆者の研究方針を規定している。動物たちは、その内部に「他者モデル」などもっていない。にも関わらず、群れという集団を維持しながら行動することができる。ここには、人間集団や、人間と機械により構成される社会を考える上で重要なヒントが隠れている。

2. 他者モデル・XAIは必要か

他者モデルとは、HAI（ヒューマンエージェントインタラクション）分野において採用されている、エージェントの内部にある他者のモデルである。現状では、単なる作業仮説や、ロボットやAIをデザインする上での工学的概念という枠を超えて、しばしば実在するものとして語られている。

筆者は著書¹⁾でも述べたように、他者モデルが実在するというということにも、エージェントの設計上それが必要であるということにも否定的である。その理由を一言でいうなら、他者とはそもそも「理解できないこと」がその重要な定義であり、一切のモデル化を拒否する存在であるからである。上述のように、動物たちは他者モデルなどもっていないにも関わらず、群れや社会行動を維持しているのだ。他者モデルとは、すべてを数値化・内部化することで理解しようという、郡司²⁾のいう意味での「人工知能」的世界観が生み出した幻にすぎないだろう。

XAI（説明可能AI）も、これと同じ世界観に基づくものであるように思われる。これはAIの処理過程を、ユー

ザである人間に理解可能なようにしようとするもので、あくまでインタフェースデザインであると考えたら有用であろう。問題なのは、このAIが出力する過程を、実際の「AIの思考過程」と混同してしまうことである。AIとは、すべてのものの価値を主体を中心とした価値観の中で評価するシステムである。AIの思考過程を人間の思考過程に置き換えるというのは、世界を構成している中心点である「私」を、ほかの「私」と機械的に入れ替えることを示している。そのようなことは原理的にできない。他者モデルもXAIも、「私」の特異性を相対化したところに生じるものであり、「世界のすべては私にとって理解可能なものである」という人工知能的原理に基づくものである。

3. 平田篤胤の天狗

理解不可能な、モデル不可能な他者というものは、かつては人間の生活の近くに息づいていた。前近代、失踪事件などが起こると「天狗の仕業」といわれたのは、天狗という世界の「外部」のエージェントにその責任を引き受けてもらうことで、社会を円滑に回していくためである³⁾。

平田篤胤が、天狗に攫われた少年に聞き取り調査を行ったことはよく知られている。当時第一の知性であった平田が、天狗や異界というものにリアリティを感じたのは、知性では到達できない領域、すなわち外部を見ようとしたためだろう。一方現代人は、平田篤胤の聞き取った天狗について「宇宙人だったのではないか」などと、自分に理解可能なものに引き寄せて解釈しようとする。これこそが人工知能的態度であり、他者モデルやXAIに拘る研究者と同じものである。

(2023年4月14日受付)

参 考 文 献

- 1) 松井哲也：ロボット工学者が考える『嫌なロボット』の作り方，青土社 (2022)
- 2) 郡司ベギオ幸夫：天然知能，講談社 (2019)
- 3) 小松和彦：神隠し，弘文堂 (1991)

.....
[著 者 紹 介]

まつ い てつ や 君
松 井 哲 也 君

2013 年神戸大学大学院理学研究科卒業。2021 年より大阪工業大学ロボティクス&デザイン学部講師。HAI 研究に従事 (理学博士)。

E-mail: tetsuya.matsui@oit.ac.jp

所属：大阪工業大学 大阪府大阪市北区茶屋町 1-45
.....