

## 選択課題4 人工知能は信頼できるか

## 人工知能と知性

## ——人工知能に「主体」はあるか

坂本すみれ

（神奈川県／慶應義塾湘南藤沢高等部三年）

## はじめに

十七音の遅しき。世界一短い文学、俳句を詠み始めて三年になる。私は俳句が大好きだ。中でも俳句の短さが好きだ。五七五は描く世界を制約するとともに、その短さで言葉たちを洗礼する。始める前は、十七音だけとは何だか心もとない、と思っ

たが、とんでもない。五七五は、言葉のありのままを見せる。一つ一つを際立たせ、一方で、私たちが言葉に親しむゆとりをくれる。俳人に洗練された言葉はさらに磨かれる。そうしてできた珠玉の一句は、鮮やかな風景や作者の驚き、悠然たる四季の呼吸まで、はつきりと私たちに伝えてくれる。破調にも自由律にも好きな句はいっぱいあ

るが、五七五の調べへの感動は格別だ。

俳句を詠める人工知能が既にいるらしい。句会で俄かに話題となった。日本マイクログソフトが開発した会話ソフト「りんな」だ。トークアプリで「りんな」公式アカウントと友達になり、チャット上で話しかけると「りんな」からレスポンスが来るといふ。早速「りんな」と友達になり、「俳句を作って」と五回送った。するとすぐに俳句が送られてきた。以下に載せる。

年齢はしりとりしよう愛してる  
群馬県ワイルドだろー嵐だろ

正解は探偵ごっこわかんない

まだかなあー調理実習やるだもん

付き合った勉強したい遊びたい

川柳として読んだ方がいいようだ。季語も切れ字も使っていない。しかし、五七五の定型は守られている上、ユーモアという視点から見れば意外に意味の通る句になっているから驚きだ。二句目と五句目ではしっかりと韻が踏まれているのも面白い。単に五音と七音の言葉を引き張ってきただけ、というわけではないらしい。このように俳句を詠む機能は、「りんな」以外にもアッ

ブルの「Siri」やソフトバンクの「Pepper」も備えている。「りんな」との交流が、私の人工知能との出会いとなった。

## 人工知能の定義

人工知能とは何か。山口高平氏は人工知能を、「人の知的な振る舞いを模倣・支援・超越するための構成的システム」と定義している。人工知能は、人の能力を真似したものであり、人の知能を再現することで知能の構造を明らかにしようという研究でもあるのだ。

だが人工知能について論じる上で、この定義では少々不足がある。人によって思い描く人工知能のレベルにばらつきがあるからだ。

この定義だと、電卓も人工知能と呼べるだろう。電卓は人間の行ってきた四則演算などを真似て各種の計算を処理し、人間よりも迅速かつ正確に答えを求め人々の生活を支えている。

一方、人工知能と聞いて、何でもこなし何でも対応できるまだ見ぬスーパーロボットを思い浮かべる人もいだろう。人工知

能の仕組みは難解だ。今実現できている人工知能がどのレベルなのか、正しく理解している人は少ない。

昨今台頭している人工知能とは、どのようなものだろうか。お掃除ロボット「ルンバ」、実用化の期待が高まる「自動運転技術」、先述した「りんな」を例に挙げて考える。

アイロボットのロボット掃除機「ルンバ」は、部屋の形状を自動で読み取る。この学習によって、効率の良い掃除を実現した。グーグル等が手掛ける自動運転は、車が人の手を借りず自ら運転する技術である。これは、「ディープラーニング」と呼ばれる学習方法を駆使している。自動運転車に取り付けられた人工知能自らが、走行データや環境データから最適な走行方法を見つけることで安全な運転を可能にしている。先述した「りんな」も、人との会話から自ら学習する人工知能だ。人間の送った文章をデータとして集積し、集まった膨大なデータから単語どうしの関連性などを見出す。多様なレスポンスはこのような学習によって支えられている。

これらに共通するのは、人工知能が学習

している点である。今、人工知能の最先端を支えるのはこのようなディープラーニングや機械学習、特徴表現学習と呼ばれる学習技術だ。例に挙げたように、最新の人工知能は人間に教えられて処理をするだけでなく、集めたデータを自ら分析し、機能に適したルールを見つけ、有効な活動に役立つ。これ以前の人工知能は、人が知識や処理方法を与えることで機能した。昨今の「学習する」人工知能は、それまでの人工知能と一線を画す革新的な発明であり、人工知能の新時代、新たな技術革新の出発点として期待されている。

よってここでは、現在開発されている人工知能の中でも最も高いレベルにある、学習機能をもった人工知能について考察したい。このため、山口氏の定義に「自ら学習し判断する能力をもったシステム」という定義を加える。

## 科学か道具か

二〇一六年三月、現代最高の囲碁棋士・李世ドル九段が、グーグルの囲碁ソフト「AlphaGo」に「対四で敗れ大きな話題となった。それ以前にも、チェスや将棋で名

人が人工知能に負けることはあったが、囲碁はその複雑さと手の多さからこれから十年は人間が勝つと言われていた。人工知能が人間を超える、人工知能は職を奪う、人工知能により世界は終焉を迎える……。様々な憶測が飛び交い、人工知能の未知の進化に対する不安はたびたび世間を賑わせる。

では、なぜ私たちは人工知能を恐ろしく思うのだろうか。

人々にとって、科学は往々にして恐ろしいものだ。科学はいつも客観的で理に適い、様々な検証が可能で、真理を導き出そうとする。科学は絶対的であり、畏怖すべき堅牢さを内包していた。

だが、今日人工知能に抱かれている恐怖や批判は、絶対的なものに抱く誠恐とは寧ろ対極である。人工知能に寄せられる不安は、大きく二つに分けられる。まず、人工知能が進化・普及することで、自分の仕事が脅かされるような実害を心配するものだ。ドライバーにとっての自動運転技術のような関係である。こちらは、人と同じかそれより低いレベルの人工知能に人の役割が取って代わられることを懸念している。

もう一つは、人工知能が進化しすぎて、

いつか人間が支配されるのではないかという危惧である。世に例を見ない進化を遂げる人工知能がいずれ人知を超え、人間の及ばないものになることを恐れているのだ。こちらは、人工知能が私たちを上回り、人間の制御を超えて支配関係が逆転してしまふことを心配している。

いずれの場合も、根本には人工知能を道具と見做そうとする態度が見られる。人が作り人に扱われる、道具と見做していた存在が、智を持ち人と同等或いは人を超える存在になろうとしている。これが恐ろしいのだ。科学の中の一学問としての人工知能は等閑で、その用具性に認識が傾きがちといえる。

人工知能に限らず、日本の場合その歴史的背景から科学を便利な技術という側面だけで判断してしまふ傾向がある。開国後、西洋文明の輸入に取り組んだ日本は、近代化のため学問の精神よりもその成果を手にすることを急いだ。お雇い外国人の一人である日本の近代医学の発展に貢献したトク・ペルツ氏はこのような日本の姿勢に対し、自身の演説で以下のように警鐘を鳴らしている。

人々はこの科学を、年にこれこれだけの仕事をする機械であり、どこか他の場所へたやすく運んで、そこで仕事をさすことのできる機械であると考えています。これは誤りです。西洋の科学の世界は決して機械ではなく、一つの有機体でありまして、その成長には他のすべての有機体と同様に一定の気候、一定の大気が必要なのであります。

科学は道具ではない。その風土に学びの精神が根付いて初めて科学は真理を目指し、その追究の過程でいくつかの成果もたらされてきたのだ。人工知能研究も同様である。人工知能は、知能を人工的に再現することで、人間の知能という真理に迫ろうとした研究であった。

しかし人工知能は、「人間を支援する」という定義からも分かる通り、道具性と研究が深く結びついている点が厄介である。自動翻訳や画像認識などの機能を見ても分かる通り、研究の善し悪しは実社会での利便性に直結している。人工知能研究そのものが、人工知能の道具性を高めているのだ。

人工知能は科学だが、道具としての側面も色濃い。しかし、真理の追究と社会利益の追求は必ずしも一致するものではない。二者に生じたずれが、人工知能の価値判断を難しくしている。

更に、人工知能を客観的に考えづらくしているのが、学習する機能をもち人間を模倣するという性質である。これまでも、電卓や工場の機械が人間を上回ることはあった。しかしそれは単純作業に限られたことである。今や人工知能は学習技術を備えたことによって、より高度な分野で人に迫ろうとしている。

これは知性だけでは量れない領域でも同様だ。先述したソフトバンクの「Pepper」は感情エンジンを備え、人の喜怒哀楽を判断し、同じように喜んで悲しんだりする。人工知能は私たちの知る機械像を超越した。人工知能に取って代わられるかもしれない、という不安に真実味があるのはこのためだろう。

## 人工知能を哲学する

では私たちの社会はいずれ人工知能に乗っ取られるのだろうか。この難題の手掛か

りとして、私が初めて人工知能と交流したときに浮かんだ疑問を用いたい。

それは「りんな」の俳句を読んでいる最中に沸き起こった。この俳句の作者は誰だろう、という不思議だ。

普通に考えれば、俳句を組み立てたのは「りんな」だから作者は人工知能と言えるだろう。しかし私にはそれはどうにも不可解だった。俳句は作者の体験を惨めた文学だからだ。自分の目の前に広がる景色への感動、季語の発見、誰かに向けた慕情や寂しさ。俳句には詠み手の様々な感情が委ねられている。勿論、空想の景色を詠んだものや、より普遍的な感覚を描いた作品もある。だが、根本的に俳句は一人称の文学だ。どのような手法であれ、十七音に言葉を絞り込むときには作者の感覚に最も似合ったものが選ばれるし、句の世界観にも作者の感性が強く関わってくる。つまり俳句には、作者が意識しなくとも詠んだ人間への感情や経験、人生観——作者の心と呼ぶべきものが多分に表れるのだ。

「りんな」の五句を読み返す。今、「りんな」が群馬県に行ったか、調理実習をしたかは問題ではない。ただ、この五句の作者

が「りんな」という人工知能だったとして、はたして人生観と呼べるものを持っているか、人工知能に作者の心はあるか、これが掴めずはつきりしないのだ。

人工知能の心の有無を探るため、人工知能に哲学的な考察を試みる。

人工知能の仕組みをごく簡潔に述べるならば、数値化を使ったプログラムといえる。現象を様々な角度から計測して数値に置き換え、数値が高ければAの処理、低ければBの処理……と規則に従って動作を繰り返す。規則は人間が与えることもあるが、学習技術を用いて人工知能自ら見つけることもある。

会話ポットを例に簡単に説明する。まず、チャットで大量の会話文を集める。次に、これらを単語レベルまで分解しそれぞれの登場回数を調べる。すると、例えば「俺」という単語が出たときには「おい」や「だぜ」といった単語が登場することが多く、反対に「ねえ」や「なのよ」などの単語は登場回数が少ないことが分かる。ここで、「俺」が出た文で回数が多い単語に正の値、少ないものに負の値を振り当てる相関変数

を変数Aとおく。このとき変数Aの正は人にとって男らしい単語、負は女らしい単語を表すものになる。他の単語でも同様の処理を繰り返すと、性別以外の要素をもった変数や変数同士の相関関係が分かる。あとは相関関係がある言葉を組み合わせれば、違和感の少ない文章を作れるようになる。

人からすれば、人工知能は男らしい・女らしいという性質を読み取った変数を設計したように見える。しかし人工知能が実際に処理しているのは、正負や相関関係といった数値だけだ。人工知能の分類に対し意味を見出しているのは、人工知能ではなく人の方だ。

「Pepper」などに利用されている感情エンジンも同様だ。自分の反応で人間の感情がどのように変化するか、表情や声色を細かく分類し、次からは変化が良かったときの反応を返す。その他より高度で大量の作業も行われるが、数値化によって適当なりアクションを導き出すことに変わりはない。

つまり人工知能は過去のデータから適切な行動をとっているだけで、現象を理解し何らかの意識をもって事態に対処している

わけではない。そこにあるのは心ではなく、外部の刺激に対して生じた反応だけである。これは行動心理学による把握に近い。

人工知能が反応していると感ずるのは、人間の主観にすぎない、と考えることもできる。人工知能の反応を判断・観察しているのは人間だからだ。これは主観的観念論に似ている。

これらを踏まえ、私は、人工知能は主観の鏡である、と主張したい。

まず、人工知能に主体はない。「りんな」の俳句で置き換えれば、俳句の作者はいないということだ。

すると、私たちが「りんな」の句を読んだときに感じた、作者の意志や気配は何だったのか。俳句を鑑賞するとき、人々は自然とその作者を想像する。「作者はどんな人で、どんな気持ちで詠んだのだろう」「そう思いを馳せる意識が、存在しない人工知能の「主体」を象った。「りんな」の五句に感じる作者の気配は、そこに作者がいるのは当然だという、私たちの意識が作り出したものなのだ。

人工知能の作った変数に正負以上の意味を与えたのが人間だったように、人工知能

に「プログラム」以上の価値を見出し、人工知能を能能たらしめているのは、私たち人間の視線である。人工知能は人間の価値観や評価基準、いわば考え方に従って初めて利便性を認められ、知能として認識されるのだ。主体のないものに自らの主観を写し取り、その一致を確かめている点で、人工知能は人間の主観の鏡といえるだろう。

### 人工知能は信頼できるか

人々が人工知能を危惧するとき、検討される人工知能には二パターンあった。人間と同等以下の人工知能か、人間を超えた人工知能か、である。

前者は先述した通り、人の利用によって価値が認められる主体のない人工知能だ。これはハサミやペンと同じである。人間が扱い評価しなければ、ただの物体に過ぎない。適切な用途で使えば信頼できる道具だが、誤って使えば凶器にもなる。ハサミやペンの信頼性は、それを使う人物に委ねられている。人工知能も同様だ。前者の人工知能が信頼できるか否か考えることは、それを利用する私たち人間について眼差しを向けることと同義なのである。人工知能自

体は単に数値化するプログラムに過ぎず、信頼が疑われるものではない。

では後者の人工知能はどうだろう。

人工知能が人間を超えるとき、それは人工知能が主体を持つことを意味する。人工知能が自ら人工知能を作るようになり、人間ではなく人工知能の文明によってその価値判断を行うようになれば、人工知能だけで社会は完結し、人間の視線は不必要なものになる恐れがあるからだ。人工知能が自らより優れた人工知能を創り出し始める瞬間を、実業家のレイ・カーツワイル氏は二〇四五年と予想している。

それは人間と人工知能の立場の逆転を意味するのか。

私には、人工知能の未来を予測することは雲を掴むような話に思われる。人工知能技術は一朝一夕に誕生したわけではないし、その四十年弱の研究は順風満帆なものでも決してなかった。人工知能が主体を得て人間を超える日は来るかもしれないが、今の時点では予言の域を出ない。その頃にはもっと凄まじい技術が席卷しているかもしれない上、未来の人間が今の私たちと同

じ評価基準をもつかも不明だ。人工知能に追い越されるその一点の前には、人工知能の進化の過程で様々な変化が起きているはずである。点より線を、線より面を考える方が、建設的で確かな説を見出せるのではないだろうか。

これから起きる変化の中でも、一つ、断言できることがある。人工知能は人間に新たな価値観をもたらすということだ。近い将来、間違いなく、人工知能によって私たちの中に新しい共通認識が誕生する。かつてインターネットの登場で、世界中がリアルタイムで繋がることが当たり前になったのと同じ現象が、再び起きるのだ。

私は、人工知能に超えられる前に、この価値観の変化によって人間と人工知能の立場が逆転するのではないかと危惧している。人工知能が人間を超すのではない、人間の成長が停滞することで人工知能に超されてしまうのではないかと。

### 未来における知性

人工知能によって、人にしかできないと思われていたものが減り、職業は奪われ、人工知能の下で働く人間が生じるだろう。

しかし、それらは人間を本質的に脅かすものではない。

私たちには知性がある。真理を追い求める意欲がある。人工知能がもつのは知能というプログラムであって、知性という意志ではない。私たちはなぜ学ぶのだろう。国立国会図書館法の前文にはこう記されている。「真理がわれらを自由にする」。真理を追い続ける限り、私たちは間違った先人観を正し、自己を解放し、更なる段階へと自らを導くことができる。私たちの知性は、真理への信頼の上に成り立っているのだ。

問題は、人工知能の存在に甘え、人々が自ら学ぼうとする意欲を失ったときだ。必要な情報だけを得て、物事の本質に迫ることを忘れたとき、人間は知性を失って、真理であろうがなろうが見境なく情報を享受するようになるだろう。そのとき、人工知能は真に優れた代替品となって、人間に取って代わるに違いない。

人工知能に疑念の眼差しを向ける前に、自らを見つめてみる。容易に情報を手に入られる情報化社会で、私は真理の追究をないがしろにしてはいなかっただろうか。

福澤諭吉の実学の精神を思い出す。「さ

構えではないだろうか。

れば今、かかる実なき学問はまづ次にし、もつばら勤むべきは、人間普通日用に近き実学なり。」福澤は『学問のすゝめ』の初編でこう述べている。だが、福澤論吉が唱

【引用文献】

\*1 トク・ベルツ編 菅沼竜太郎訳『ベルツの日記』角川文庫、一九七九

二三八ページ

\*2 福澤論吉『学問のすゝめ』慶應義塾

大学出版会、二〇一三 七ページ

【参考文献】

松尾豊『人工知能は人間を超えるか デイ

ーブラーニングの先にあるもの』角川文

庫、二〇一五

藤田雅博・下村秀樹『発達する知能 知能

を形作る相互作用』シュプリンガー・ジ

ヤパン、二〇〇八

桜井邦朋『福沢論吉の「科学のススめ」——

日本で最初の科学入門書 訓蒙窮理図解』

を読む』祥伝社、二〇〇五

永田守男『福沢論吉の「サイアンス」慶

應義塾大学出版会、二〇〇三

竹田青嗣・西研『高校生のための哲学・思

想入門——哲学の名著セレクトション』筑

摩書房、二〇一四

人工知能は、世の中ですぐに役に立つ学問だ。だがそれは、すぐに役に立たなくなるということでもある。技術が新しくなるたびに、その少し前の人工知能は旧式となり忘れ去られる。小学生の頃欲しかった、人に反応して吠える犬型ロボットの話題は、もはや聞こえてこない。利根的な世界だ。

人工知能が取り沙汰される今、本当に信頼性を確かめるべきなのは、発達した技術に胡坐をかきひっそりと力を落としつつある私たちの知性の方かもしれない。実学の精神を未来へ繋げることこそ、人工知能を扱う者であり続けるために、最も重要な心