

# 人工知能は知識を持てるか？

久木田水生<sup>\*†</sup> 戸田山和久<sup>\*</sup>

2019 年 4 月 21 日  
応用哲学会第 11 回大会  
京都大学

## 1 はじめに

2011 年、IBM の開発した人工知能システム Watson が、アメリカのクイズ番組「ジェパディ！」に出場して人間のチャンピオンと戦った。Watson は膨大な事実のデータベースを持ち、例えば「スティーブンソンの『宝島』の悪役は」といった問題が与えられるや否や、データベースを検索し、人間よりも素早く回答を返すことができた。結果として Watson は圧倒的な差をつけてチャンピオンを破った<sup>\*1</sup>。

私たちは通常、クイズに強い人間を「物知り」、つまり「知識」を豊富に持っている人間とみなしている。では Watson も同様に「物知り」と言えるのだろうか？ しかし私たちは通常、大量の情報が蓄えられたコンピュータに対して、それが何かを「知っている」というような記述の仕方はしない。もしコンピュータが何かを「知っている」ということができるならば、多くの情報が印刷された百科事典を「物知り」と呼ぶことも出来るのではないだろうか。それはかなり馬鹿げたことであるように思われる。もし Watson が通常のコンピュータと本質的に違いがないのであれば、それが何も知っているということではできないだろう。

1950 年にアラン・チューリングは「計算機械と知能」と題された論文において、「機械は考えることができるか？」という問いを提起し、その中で後に「チューリングテスト」として知られるようになる、機械が考えているかどうかを判定するためのテストを提案した (Turing [8])。彼の提案はその後、賛否両論の激しい論争を引き起こしたが、そのおかげでそもそも「思考とは何か？」という問題についての私たちの理解は以前よりもずっと深まった。

このように人工知能は、人間の様々な知的能力（計算、推論、計画、言語処理、パターン認識、身体制御、感情などなど）をシミュレートすることを試みることで、その知的能力がそもそもどのようなものであるかについての私たちの理解を深めてくれる。学問分野としての人工知能の意義の一つはここにある。本稿では「人工知能は知ることができるか」という問いにアプローチすることで、「知識とは何か？」という古くからの哲学の問題について考える新しい視点を提供することを試みる。

---

<sup>\*</sup> 名古屋大学情報学研究科

<sup>†</sup> minao.kukita@gmail.com

<sup>\*1</sup> <https://www.ibm.com/watson/jp-ja/quiz/index.html>

## 2 人工知能と心の哲学

「人工知能 artificial intelligence, AI」は、コンピューターによって「知的」な振る舞いを実現することを目指す分野として、20 世紀の半ばに誕生した。しかしながら「知的」な振る舞いとはどのようなものかということは必ずしもはっきりしていない。そのため人工知能に従事している研究者たちが目指しているものは多種多様であり、時代によっても大きく変化してきている。

初期の人工知能は主に論理的な推論能力に焦点を置いた研究が多かった。論理的な推論とは例えば「すべての X は Y である」と「a は X である」という二つの文から「a は Y である」という文を導き出すような推論である。19 世紀末から 20 世紀の初頭にかけて発展した記号論理学は、少数の規則を組み合わせることで、あらゆる論理的な推論が実現できることを示した。この成果をコンピューターに応用して、コンピューターによって自動的に論理的推論を実行させたのが、初期の人工知能の一つの顕著な達成であった。初期の人工知能の発展に大きく貢献し、その功績によってチューリング賞を受賞したアレン・ニューウェルとハーバート・サイモンは「例えばコンピューターのような、物理的記号システムは知的な行動のための必要にして十分な手段を持っている」とする「物理的記号システム仮説」を提唱した (Newell and Simon [6])。このような方法論は「論理的アプローチ」と呼ばれている。

しかしながら記号的アプローチでは人間の知的振る舞いの多くにとって十分ではないということが次第に認識されるようになった。例えばチェスは高度に知的なゲームと考えられるが、論理的に考えるだけでは強いチェスプログラムを作ることは困難であった。強いチェスプレイヤーは、論理だけではなく経験に基づく勘にも大いに頼っている。そこである分野の専門家、エキスパートの「勘」を明示的に規則化して、それを判断の基礎に据えるというアプローチが探求されるようになった。このようなアプローチは「知識アプローチ」と呼ばれ、このアプローチで作られた人工知能は「エキスパート・システム」と呼ばれた。知識アプローチはいくつかの分野においては一定の成功をおさめ、実用的なエキスパート・システムも作られた。しかしエキスパートの知識を人工知能に与えられる規則のような仕方で明示化することはしばしば難しく、このアプローチの成功も限定的であった。

論理的アプローチ、知識アプローチが苦手なことに実世界の環境の中で適切に行動するということがある。例えば部屋の中で障害物を避けて歩き回るといったような機械を論理的アプローチでつくり上げるのは極めて困難であった。しかしながら昆虫のような単純な生物でも、環境の中で適切に行動できる。人間を含めた生物は、環境と身体の力学的な相互作用を利用することで環境の中で適切な行動をしているのである。そこでそのような環境との相互作用を利用するアプローチが提案された。これは「身体化アプローチ」と呼ばれる。

現在の人工知能ブームのきっかけになったのは「ニューラルネットワーク」という脳の中のニューロンのネットワークを模したモデルである。脳の中では単純なルールに従って信号のやり取りをするニューロンが膨大に集まって、その集積によって複雑な認知や思考を実現している。そこでは言語的に表現されたルールはなく、ニューロン全体が作るネットワークの構造が、知的な振る舞いを創発させているのである。

このように人工知能という分野は必ずしも、明確に定義された目標を持ってそれを実現するために研究開発をしてきたのではない。むしろ人工知能は、「知的な振る舞い」という多分に曖昧かつ多様な現象がいかなるメカニズムによって実現されているのかについて様々な仮説を立てながら、その仮説に基づいたシステムを構築し、そのことによって仮説の正しさを確かめようとしてきた。論理的 AI、エキスパート・システム、身体化された知能、ニューラルネットワークなどはそれぞれが「知的な振る舞い」とはどのようなものであり、それはどのように実現されるか、という問いに対する一つの仮説を表している。

哲学もまた古くから人間の知能あるいは精神がどのようなものであるかということを探求し、様々な仮説を立ててきた分野である。しかし近年に至るまで哲学はその思弁的な仮説をテストするための実証的な手段を持っていなかった。20 世紀に入り、神経科学や心理学、計算機科学、あるいは生物学は哲学者の考えた仮説を補強する、あるいは棄却する実証的な根拠を提供した。それだけではなく、これらの分野は哲学者が心について新しい仮説を立てる際に大きなインスピレーションを与えた。20 世紀の哲学はこれらの分野の圧倒的な影響力の下に、「心の哲学」という新しいサブジャンルを発展させてきた。本発表が目指すところも同様であるが、本発表では特に、人工知能の発展に照らして「知識」という概念について再検討しようと思う。

### 3 哲学の二つのアプローチ

哲学者が「知識とは何だろうか」というような問題に取り組む際に、大きく異なる二つのアプローチがある。一つは「概念分析」と呼ばれるアプローチである。もう一つは近年「概念工学」と呼ばれるようになったアプローチである（Cf. 戸田山・唐沢 [12]）。

概念分析を行う際、哲学者は、問題にしている概念には決まった正しい使用法があり、私たちは（少なくともある程度の知性や常識を持った人間であれば）それを正しく使用できており、しかしながらその正しい使用の条件について明示的に意識していない、という前提に立っている。その上で、様々な正しい使用と正しくない使用を見比べて、その概念が正しく使用されているための必要十分条件を特定しようとするのである。

知識についてであれば、例えば「太郎は次郎の血液型が A 型であることを知っている」という言明は、どういう場合に正しく、どういう場合に正しくないかはすでに決まっている、というのが概念分析をするための前提である。その上で、私たちの常識的な判断\*2に基づいて、この種の言明が正しいための必要条件を特定するのが哲学者の仕事、ということになる。

古典的には「太郎は次郎の血液型が A 型であることを知っている」という言明が真になるためには次の三つの条件が必要十分であると考えられてきた。

- (1) 次郎の血液型は実際に A 型である。
- (2) 太郎は次郎の血液型が A 型だと思っている。
- (3) 太郎はそうのように考える正当な根拠を持っている。

というのも、もし次郎の血液型が A 型以外であれば上の「太郎は次郎の血液型が A 型であることを知っている」という言明——以下、これを S と呼ぶことにする——は真ではないと通常は考えられるだろう。また太郎が、次郎の血液型は B 型だと思っているとしたら、その時もやはり言明 S は間違っているように思われる。さらには太郎が仮に「次郎は几帳面な性格だから A 型に違いない」といういい加減な理由で A 型だと考えるようになったとしたら、たとえ (1) と (2) の条件が成り立っていたとしても、私たちは言明 S が正しいとは考えないだろう。逆に (1)-(3) が成り立っていれば私たちは言明 S が正しいと思うだろう。それゆえこの三つの条件は言明 S が成り立つための必要十分条件だ、という訳である。ここから伝統的に「知識とは正当化された真なる信念である」と言われてきた\*3。これが概念分析の一つの典型的な例である。

概念分析とは異なるアプローチとして、近年「概念工学」というアプローチが提唱されている。このアプローチを提唱する哲学者は、概念には固定された使用法、使用の条件が定まっているわけではなく、社会の変

---

\*2 実際には特定のグループに属する人たちの間でのみ共有された判断、あるいは個人的な判断であることも多い。

\*3 知識のこの特徴づけについては、これだけでは十分ではないという反論も出されているが、その点についてはここでは踏み込まない。

化や科学技術の進歩に従って創造、修正される、と考える。そして哲学者の仕事は、既存の使用法を分析して、正しい使用のための条件を明示化すること（だけ）ではなく、その時代や社会にあった概念の使用法、使用の条件を考え提案することだ、というのである。

例えば「権利」という概念について考えてみよう。この概念の適切な使用の条件は時代によって大きく変化してきた。18世紀末、イギリスの思想家メアリ・ウルストンクラフトが『女性の権利の擁護』（Wollstonecraft [10]）という著作を書いたときには、「女性は男性と同等の権利を有する」という言明は正しいもの、すなわち権利という概念を正しく使用したものとは見なされていなかった。ウルストンクラフトは女性も男性と同等の権利を持つと主張したのであるが、これは概念工学の言葉で表現すれば、それまでの権利概念の使用の慣習に異議を唱え、新しい使用の条件を提案したのだとすることができる。権利という概念はその後「動物の権利」といったようにその主体の範囲を拡大され、最近ではデヴィッド・J・ガンケル (Gunkel [2, 3]) のように「ロボットの権利」を考察している哲学者もいる。また社会の変化や科学技術の発展に伴い、「忘れられる権利」\*4、「人工知能の決定に服さない権利」\*5のような新しい権利が提唱されている。

このように概念というものは時代によって変化するものであり、かつそれは目的や意図をもって変化させられるものでもある。概念工学とは、そのような意図的な概念の修正・更新を意味するものであり、私たちはそれが哲学者の仕事の一部であると考えている。

## 4 知識の越境

アンディー・クラーク (Clark [1]) によれば、人間の最も際立った特徴は、自らの作り出した道具と環境によって知能\*6を拡張することができることである。このように主張する際に、クラークが用いる印象的な例が書き言葉である。書き言葉（あるいはそれに類するもの）の発明によって人間は生身の脳に蓄えておくことができない大量の情報を保存し、好きな時に参照することが可能になった。このように生物学的身体の外部的環境の中に保存された情報も、私たちの生物学的脳の中に記録された情報も、私たちが生存や繁殖のための課題を解決するために有用であるという意味では私たちの心・知能の一部をなしているものであり、それゆえ私たちの心は生物学的身体を超えて、道具や環境にまで広がっている。これがクラークの提唱する「拡張された心」仮説である。

「拡張された心」仮説に従えば、私たちの知識は私たちの脳の中に閉じ込められたものではない。実際に私たちは「誰々のメールアドレスを知ってる？」と尋ねられたとき、もしそのアドレスが自分のコンピューターや携帯電話の中に登録されていれば、自分が記憶していなくても「知っている」と答える。戸田山 [11] も同様に、知識が個人の頭の中に保持されている必要はない、と主張する。その際に戸田山は特に科学的知識の社会性に訴える。多くの科学者によって遂行されたプロジェクトにおいては、そのアウトプットである結論を指示する数多くの証拠は、そこに携わった研究者たちの間で分散されて持たれている。おそらくその中の誰一人として、その結論を導き出す根拠のすべてを自分で確かめてはいない。彼らは共同研究者が確かめてくれたことを信頼して、その上に結論を導いている。むしろ科学という営みそのものが、多くの科学者によるビッグブ

---

\*4 清水陽平「忘れられる権利のいま」、<http://synodos.jp/society/11698>

\*5 Article 22 EU GDPR “Automated individual decision-making, including profiling”, <http://www.privacy-regulation.eu/en/article-22-automated-individual-decision-making-including-profiling-GDPR.htm>

\*6 クラークは「知能」を、生存や繁殖などの目標を達成する際に解決しなければならない問題に取り組むための能力と捉えている。クラークが著書の中で「心 mind」という言葉を使う時も同じように問題解決の能力を指している。この定義に従えば、人間が数学の命題を証明する能力も、コオロギのメスがオスの鳴き声を頼りにオスの居場所に近づいていく能力も、どちらも知能である。

ロジクトのようなものであるとも言える。また戸田山はクラークと同様にコンピューターなどに保持された情報もまた私たちの知識を構成する要素であると考え。さらにそこから一歩進んで、戸田山は知識の保持のみならず、知識の発見もまた人間ではない人工物、すなわち人工知能によって行われる可能性を示唆する。

## 5 Watson は物知りなのか？

ここで最初の問題に戻ろう。Watson は確かに多くの情報を持ち、質問に合わせて求められる情報を即座に提示することができた。では Watson は例えば「スティーブンソンの『宝島』の悪役はジョン・シルバーである」ということを知っていることになるのだろうか？ もし古典的な知識の特徴づけ（「正当化された真なる信念」）に従えば答えは否だろう。信念という言葉の通常の意味から考えれば、それは心の中に生じる思考であって、その主体が事実と合致していると受け入れているものだろう。ではワトソンは思考を持っているのか？

この問いに答えるためには、そもそも機械が思考するということはどういうことなのかを考えなければいけない。この問いに関して最も有名な提案はチューリングによる、「人間と区別がつかないくらい自然に会話ができる機械は考えているとみなして良い」という提案である。これは中々ラジカルな提案で、もちろん多くの批判にさらされた。しかし仮にチューリングの提案を受け入れたとしても Watson はおそらくこの基準をクリアしない。というのも Watson は「それについて書かれた文献が存在しない、日常的な知識の解釈を要求されるような問題」を解くのは苦手だと評価されているからだ<sup>\*7</sup>。

ではもし Watson に十分な常識的知識のデータを与えて、どんな問いかけに対しても普通の人間と同じくらい自然に応答できるようになったらどうだろう。それでもやはり Watson が思考しているとは言えない、という結論が哲学者ジョン・サール [7] のチューリングに対する批判から導かれる。サールの反論は一言で言えば、コンピューターは自分が操作している記号の意味を理解していない、それゆえ考えているとは言えない、というものである。確かに Watson は「スティーブンソンの『宝島』の悪役は？」と聞かれて「ジョン・シルバー」と答えることができる。しかし Watson は「悪役」などの言葉の意味を理解していない。もちろん「悪役」の定義を別の言葉を使って述べるようにプログラムすることはできるだろうが、その別な言葉についても Watson が意味を理解していないことには変わらない。どこまで行っても記号と記号の間を堂々巡りするだけで、Watson にとって記号が現実の世界の何かを意味することはないのである。この問題は人工知能の分野では「記号接地問題 the symbol grounding problem」(Harnad [4]) と呼ばれている。

サールのチューリング批判にも様々な反論があるが、無意味な記号（現実の世界に接地していない記号）の操作では思考と呼ぶのに十分ではないという点は射ていると思う。だとすれば Watson は思考や信念を持つてはいない。それゆえ古典的な定義では知識も持ってはいないということになる。

信念については、行動主義的に解釈することもできる。すなわち信念とは、その主体の持つ欲求や必要や選好と結びついて何らかの行動を生み出す傾向性にとらえるのである。しかしこの解釈をとると、なおさら人工知能が信念を持つてるとはいいがたい。というのもすくなくとも現時点での人工知能は欲求や必要や選好を持つてはいないからである。

クラーク [1] の「拡張された心」仮説によれば、Watson はそれを利用する人間の知能の一部を担うものであるが、Watson それ自体は知識を有する主体ではない。一方、戸田山 [11] の社会化された認識論によれば、Watson のような人工知能は社会という集合的な知識主体の一部を構成している。しかし Watson 自体が知識を所有する主体となりうるかどうかについては、戸田山の理論からは明確なことは言えない。社会化された認

---

<sup>\*7</sup> <https://www.ibm.com/watson/jp-ja/quiz/index.html>

識論は、社会という集合体が知識の担い手になりうるということを主張するもので、個人や個々のデバイスについて何かを述べるものではない。

## 6 知識概念を変更することの意義

古典的な知識の定義は古代ギリシャ時代、プラトンの著作にまでさかのぼる。クラーク、戸田山の議論は、人間の認知や知識がますます技術や社会に依存するようになり、生物学的な個人の中に閉じ込められた知識だけを考えていることが不適切になった状況に対応して作られてきたものである。このような知識概念の改定の意義はどこにあるのだろうか。かつて天文学者が「惑星」という言葉の定義を変更して、それまで太陽系の惑星とされていた冥王星が、それ以降は惑星とは呼ばれなくなった。しかしこの変更は天文学者や天文愛好家以外の人々にとってはそれほど影響のないものであろう。しかし知識概念の変更はそれよりも大きな影響がある。というのも知識は私たちが生きていくうえで必要なものであり、知識を追求し、知識を利用することは私たちにとって重要な活動だからである。そのため知識をめぐる私たちの実践を正しく理解することを助けるように知識概念を修正していくことは有用である。

このことは「知識がある」という属性の帰属が賞賛を伴う評価語であるということを考えるとなおさら重要である。「知識がある」ということが、その人の生物学的脳の中に記憶されているということだけを意味するのであれば、デバイスを巧みに操作して情報を効率よく検索、利用できる人間よりも、できるだけ多くの知識を暗記している人間の方が豊富に知識を持っているということになり、より賞賛に値するとみなされるかもしれない。もちろん生物学的脳の中に多くの情報を蓄えておくことの素晴らしさは賞賛されるべきである。同じ一つの脳の中にある複数の情報やアイディアが結びつくことでしか生まれない新しいアイディアもあるだろう。しかしながらテクノロジーを使いこなして情報を活用できるということもまた、それとは別の評価軸で賞賛されてしかるべきである。

ワードプロセッサが普及しだした当初、ワープロに依存しすぎることによって漢字を覚えなないといったことが懸念された。しかし私たちの認知能力は生物学的身体とテクノロジーからなる複合体に帰属するのだとするクラークの考え方に従えば、ワードプロセッサを使うことで私たちの漢字使用能力は衰えるどころかはるかに向上しているのである。テクノロジーに頼らずに何かをする能力の価値はそれとして認められるべきだが、それに劣らず、新しいテクノロジーをより使い勝手がよくロバストなものに改良すること、そして個人個人がテクノロジーをより賢く使いこなす能力を持つことの重要性が強調されてしかるべきである。

同様に戸田山の「社会化された知識」にも重要な意義がある。現代において科学的知識は多くの科学者からなる共同体の中に分散され、互いを支え合っている。そして異なる分野の知識が思いがけない仕方で結びつくことで重要な発見につながることもある。このことを強調することは、科学的に重要な業績が特定の個人に帰されるべきではないということを教えてくれる。私たちはともすれば科学においてもヒーローを求め、特定の個人を賞賛したがる傾向がある。このような考えからは、優れた研究能力を持つ研究者に投資するのが科学を振興させる良い方針だ、という結論が導かれる。そして実際に近年の日本の科学技術政策においては、優れた研究能力を持つ（と思われる）研究者や研究機関に資金を集中させようという動きが顕著である。しかし現代の科学においては、多くの「称えられない英雄たち unsung heroes」なくして画期的な業績はあり得ない。社会化された認識論が広まることは、科学における過度の選択と集中が、科学の振興のためにはかえって逆効果であることを理解させる助けになるかもしれない。

## 7 知識概念のさらなるアップデートは有用か？

上述のように、「拡張された心」仮説においても社会化された認識論においても、人工知能が人間と独立にそれ自体として知識を持つとは言えない。しかしクラークや戸田山の議論は今から 15 年以上前に考えられたものである。この 15 年の間は ICT、特に人工知能においては飛躍的な発展があった。コンピューターはもはや単に人間が入力した情報を保存しておき、あとで簡単に取り出せるようにしておくだけのものではない。世界中の人々がスマートフォンなどのデバイスを通じて、日々インターネット上に提供している膨大なデータから、人間では見つけることのできない微かなパターンを見つけ出し、それに基づいて人々にラベル付けして、人々の属性や行動を予測することができる。このような人工知能、あるいはもっと高い自律性を持つ将来の人工知能が、人間とは独立にそれ自体として「知識を持つ」と言えるように、知識概念を更に改訂することは可能だろうか。また可能だとして、そのことには何らかの意味があるだろうか。

哲学者や人工知能研究者の中には、人工知能やロボットを擬人化して「考える」とか「感じる」といった言葉を適用することに反対する人々がいる。私も基本的には彼らに同意するが、しかし場合によっては既存の概念を人工知能にも拡張して適用することにも意味があると思う。例えば人工知能に対して「推論する」、「判断する」、「学習する」などの言葉を使うのも厳密に言えば概念の濫用かもしれない。ジェリー・カプランは、そもそも「知能」という言葉を使うことが人工知能に関する誤解と混乱を招いている、と述べる (cf. Kaplan [5])。

しかし推論や判断、学習といった言葉を人工知能に拡張することは無害であるように思われるし、また人工知能が行っていることを分かりやすく表現する便利な方法である。一方で人工知能が「気にかかる」、「欲する」などと表現することは多くの場合、実際にその人工知能が行っていることを正しく理解する助けにはならないだろう\*8。

私たちの言葉、あるいは概念は可塑的なものである。比喩的な仕方では拡張された言葉の使用法が時を経て標準的な使用法になることもある。ある言葉が特定の状況で本来の意味とは違うものとして定義され使用されるうちに、その新しい意味が本来の意味を侵食することもある。そのことはコミュニケーションをより効率的にする、あるいは対象となっている事物の理解を促進することもあれば、無用な誤解や摩擦を生じさせることもある。だからこそ概念工学に従事するものは、概念を修正することの意義について自覚的である必要がある。

では「知る」あるいは「知っている」という概念を人工知能に適用することは、人工知能が行っていることをよりよく理解する助けになるだろうか。そうではないと私たちは主張したい。人工知能による何らかの活動を、「人工知能が知識を持っている」と表現することは、現在の人工知能の能力、あるいは知識をめぐる私たちの実践を正しく理解することにはつながらない。以下でその理由を述べる。

上述したように「知識がある」という言葉は賞賛を伴う評価語である。これは、私たちにとって知識が有用なものであり、知識を獲得したりそれを人に伝えたりすることが社会にとって価値のある活動と見なされるからだろう。実際、知識は人間がこの複雑でリスクに満ちた世界の中でより適切に行動することを可能にしてくれる。しかしその一方で私たちが知識を伝達する媒体である言語は、誤った情報を伝えることもできる。そして誤った情報を伝達することは、受信者を大きなリスクにさらすことにもつながりかねない。つまり、相手に何かを伝え、それが事実であるということを請け負うということは、相手に利益や危害を与える可能性のある行為だということである。

---

\*8 あらゆる場合にこういった表現が濫用であって望ましくないと言うつもりはない。例えば人事採用に用いられた人工知能が男性に有利なバイアスを持っていたということを「男性をひいきしていた」と表現することは、人工知能に感情があると思わせるようにミスリードする書き方でなければ、必ずしも悪くはない。

知識に対して正当化が求められる理由の一つはここにある。ある人が「私は……ということを知っている」と発言したならば、その人はその内容が事実であるということを請け負ったのであり、そして聞き手がそれを信じて行動した結果に対して少なくとも部分的に責任を負うことになる。もしそれが事実でなかった場合、聞き手は発言者を責めるかもしれない。その際に、発言者がそれを事実だと思った十分な理由を提示できれば、その責任（の一部）は免除されるだろう。十分な理由を提示できなかった場合は、発言者は責任を負わなければならないだろう。このように知識は、複数の人間のコミュニケーションの中で、相手に情報を伝達する、それに対する報酬を受ける、あるいはそれに対する責任を負う、といった社会的な実践の中に組み込まれたものである。人工知能研究者のテリー・ウィノグラードと哲学者のフェルナンド・フローレスは、人工知能は社会に対するコミットメント（責任を伴う関与）を欠いているために、本当の意味で言語を理解することができない、と述べている（Winograd and Flores [9]）。同じ理由で（少なくとも現在の）人工知能は通常の意味で「知る」ことができない。それゆえに人工知能が何かを「知っている」と言えるような仕方では知識概念を変更することは、知識の持つ社会的・規範的な次元を見失わせるという理由で、有益ではない。

ただし私たちの使っている知識概念には、必ずしもこのような社会的・規範的な次元を含まないような使い方もあるように思われる。例えば学習をしたラットが「レバーを引けば餌が出てくる」ということを「知っている」というのは自然である。しかしラットはこの知識を他者に伝えることは（おそらく）ないし、そこには社会的な次元は含まれない。またミツバチは餌のある場所を「知って」おり、かつミツバチはその「知識」を他の個体に伝え、共有する。ここには社会的な次元は含まれるが、しかしおそらく責任や賞賛、コミットメントなどの規範的な次元は含まれていない。このような意味での「知識」においては、おそらく正当化は必要条件ではない。それゆえそれは哲学者が伝統的に論じてきた「知識」とは大きく異なる。このような知識概念ならば人工知能に適用するハードルはより低いかもしれないが、この点については今後の検討課題としたい。

## 8 おわりに

本稿において私たちは「人工知能は知識を持つことができるか」という問題を扱った。知識とは「正当化された真なる信念」であるという古典的な定義では人工知能が知識を持つとは言えない。そもそも人工知能が何かを「信じている」ということが言えないからである。しかしアンディ・クラークや戸田山和久は、知識主体が生物学的な個人に限定されるのは不適當であると考え、情報機器に保存されたデータもまた私たちの知識の一部を構成していると言えるように、知識の概念を変更してきた。このように、概念は科学技術や社会の発展によって変化する私たちの実践を反映するようにアップデートされるべきものである。そこで私たちは、近年の人工知能技術の発展を踏まえて、人工知能それ自体が知識の主体であると言えるように知識概念をアップデートすることが可能か、あるいは有益かということを考察した。そしてその結果として、そのような知識概念の修正は、知識のもつ社会的・規範的な次元、とりわけ知識の伝達に伴う責任という次元を見失わせることになり、有益ではないということを論じた。

しかしこの結論はあくまでの現時点での人工知能技術の水準と、知識をめぐる私たちの実践に照らしたものである。人工知能がより自律性を高めて、何らかの仕方では責任を持つ主体であると考えられるようになった時、あるいは知識の伝達において個人がコミットメントや責任を求められなくなった時、人工知能を知識の主体と見なすことができるようになるかもしれない。



## 参考文献

- [1] A. Clark. *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Oxford University Press, New York, 2003. 翻訳『生まれながらのサイボーグ心・テクノロジー・知能の未来』(呉羽真・久木田水生・西尾香苗訳, 春秋社, 2015 年). .
- [2] D. J. Gunkel. *The Machine Question: Critical Perspective on AI, Robots, and Ethics*. The MIT Press, Cambridge, 2012.
- [3] D. J. Gunkel. *Robot Rights*. The MIT Press, Cambridge, 2018.
- [4] S. Harnad. The symbol grounding problem. *Physica D*, 42:335–346, 1990.
- [5] J. Kaplan. *Artificial Intelligence: What Everyone Needs to Know*. Oxford University Press, 2016.
- [6] A. Newell and H. Simon. Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19(3):113–126, 1976.
- [7] J. R. Searle. Minds, brains and programs. *Behavioral and Brain Sciences*, 1:417–424, 1980.
- [8] A. M. Turing. Computing machinery and intelligence. In A. R. Anderson, editor, *Minds and Machines*, pages 4–30. Prentice Hall, 1964.
- [9] T. Winograd and F. Flores. *Understanding Computers and Cognition: A New Foundation for Design*. Addison-Wesley Professional, 1986. 邦訳『コンピュータと認知を理解する——人工知能の限界と新しい設計理念』、平賀譲訳、産業図書、1987。.
- [10] M. Wollstonecraft. *A Vindication of the Right of Woman*. W. W. Norton and Company, New York, 3rd edition, 2009.
- [11] 和久 戸田山. 知識の哲学. 産業図書, 東京, 2002.
- [12] 和久 戸田山 and かおり 唐沢. 〈概念工学〉宣言！：哲学×心理学による知のエンジニアリング. 名古屋大学出版会, 名古屋, 2018.