【優秀論文】

人間とAIの立場の現状と展望

～生成AIは文章を理解することができるのか～

3年8組13番　小林琉斗

Ⅰ　はじめに（アブストラクト）

AI[1]は、特に2022年から23年にかけて空前のAIブームによって、凄まじい速度で成長を遂げている。顔認証や自動運転技術に認識・識別AIが利用され、他にも、画像生成や文章生成を行う生成AI等が様々な分野で活用されてきている。しかし最近では、AIによって私たち人間の仕事は奪われる、シンギュラリティ[2]がもうすぐ訪れる、人間とAIの立場が逆転するのではないかなどの不安を掻き立てる情報だけが、メディアや専門家の発信によって独り歩きしている側面もある。人間とAI双方の関係は今後どう変化していくのだろうか。

本論文では、文章理解という観点から双方を比較していく。まず、人間と生成AIの文章理解方法の相違点を比較し（Ⅱ）、実際に人間と生成AIはどれぐらい文章を「理解」できるのかを検証する（Ⅲ）。それからAIの「理解」の現状と今後の展望について述べ（Ⅳ）、最後に双方のパワーバランスは今後どのように変わっていくのか考察していく（Ⅴ）。

Ⅱ　人間と生成AIの文章理解の方法

そもそも「理解」は、「物事の道理をさとり知ること。意味をのみこむこと。他人の気持ちや立場がよく分かること」と説明されている（広辞苑）。もし「理解」概念がこの通りなのだとすると、「他人の気持ち」など推しはかりようのないAIが本当に「理解」しているかどうか、客観的に知ることは難しい。

脳科学者の山鳥重は「言葉が分かる」ということを、「話したり、文を書いたりして運動化出来ること」と定義づけている[3]。新しい言葉や概念を理解するとき、頭の中にぼんやりしたものがあるだけでは言葉がわかったことにならない。新しく身につけた言葉や概念の形をはっきりさせることが言葉がわかる、ということである。それをもって、内容の正確な把握ができたということになる。

さらに山鳥は、本当に分かったことは、他に応用することができるとも述べている。知識の裏にある原理が理解できていれば、その知識をほかの現象にも転用することができるということだ[4]。

このように、一口に理解するといっても、その捉え方は様々である。ゆえに「理解」の本質を一言で説明することはできず、理解したという客観的な状態を証明することもできない。

このような事態に対して、教育学者のG.ウィギンズとJ.マクタイは「理解」するという概念を、説明する、解釈する、応用する、パースペクティブ[5]を持つ、共感することができる、自己認識を持つの6つの側面から捉えている[6]（以下、「理解の6側面」と表記）。ウィギンズ＆マクタイに従えば、人間は真にその物事を「理解」している時、これら6つのことが全てできるということになる。逆に言うと、人間は文章を理解する際にも、与えられた文章を6つの側面から「理解」しようとしていると考えられる。

では、人間が「文章」を理解する際、どのような構造に注目しているのだろうか。

---

[1] AI（人工知能）の定義は専門家によっても分かれるが、本論文ではAIを活用した技術のことをAIと呼ぶ。
[2] 技術的特異点。「真の意味でのAI」が自立的に自分自身より能力の高い知能を生み出す事が可能になる時点のこと。ヴァーナー・ヴィンジ氏によって提唱され、当初は2045年に到達すると予測されていたが、現在ではもっと早まると発言している人やそもそも来ないのではないかと断言する人もいる。
[3] 山鳥重（2002）『「わかる」とはどういうことか』筑摩書房、p.203-207
[4] 同前 p.208-210
[5] 将来を見通すこと。今回の場合は、批判的な目や耳を用いて、複数の視点から見たり聞いたりし全体像を見ること。
[6] G.ウィギンズ・J.マクタイ／西岡加名恵訳（2012）『理解をもたらすカリキュラム設計―「逆向き設計」の理論と方法』日本標準、p.99-102

心理学者である内田伸子は、文章理解の際、情報を受け取る側は、「一貫性」「最適性」「開放性」の3つを察知するセンサーを持っていると提唱した[7]。人間は文章に一貫性を求め、他の知識を照合したり、一度構成したものから新たな問や矛盾を発見し解消したりすることで、「理解」を深めていくことができる。特に一貫性は情報間に矛盾があることに気づき、言葉や世界に関する知識（常識）である既有知識と照合することで、不自然さを吟味し解消する[8]。この「不自然さを解消する」ことを鈴木考子[9]は「矛盾の解消」と定義づけた[10]。

対して生成AIの文章理解の方法は大きく異なる。現在の生成AIは「記号接地」を全くせずに言語を学んでいる[11]。人間は知っている言葉が指す対象を知っており、自分自身の体から、概念の様々な特徴を捉え知ることができる（記号接地をして言葉を理解できる）。それに対して生成AIは、身体を持たず接地した経験がないため、感覚に接地していない（真の意味を理解していない）記号を結びつけ、置き換えて言葉を理解しているかのように見せている。

2018年、Google社によってBERTという文章の並び方に確率を割り当てる確率モデルである「言語モデル」が発表された。その言語モデルのうち「計算量」「データ量」「モデルパラメータ数」の3要素を大規模化したものを「大規模言語モデル」（以下、LLM）と言い、代表的な例としてOpenAI社のGPTを挙げることができる[12]。つまり現在のチャットボット等の生成AIは、与えられたタスクを達成するために確率の高い言葉を次々に出力しているだけであり、真にその言語を「理解」しているわけではない。

以上のことより、人間と生成AIでは文章理解の方法が違い、生成AIは見かけ上の理解を行っているにすぎない。そのため人間が文章を「理解」するときに働く理解の6側面は、生成AIに対しては働かないのではないか。次の章ではこの点について検証していく。

Ⅲ　矛盾する情報理解の方法と能力の分析

（1）人間の場合

前述した鈴木考子は、人間が矛盾した情報を理解する方法についての研究[13]を行った。鈴木論文では①矛盾の解消をするパターンに規則性があるのか、②暗黙の条件を表面化すれば、特殊状況を考え出すことができるのか、の2つに着目し、それぞれに対して調査を行った研究成果が紹介されている。

まずは「矛盾の解消」についてである。矛盾の解消とは情報間の矛盾を一貫性・整合性のある文章に再構築し、指摘することである。鈴木は例として久保ゆかり[14]が行った研究[15]を使って説明している。

> 被験者に提示される情報は、「マアちゃんはおやつにアイスをもらった」という言語情報と、その時のマアちゃんの表情を描いた絵（普通に見ると「悲しそうな顔である」と判断されるような表情が描かれている）の2つである。仮に前者を情報aとし後者を情報bとしよう。情報aとbが矛盾していると判断し得るのは、情報aを「マアちゃんはうれしい」と解釈し、情報bを「マアちゃんはうれしくない」と解釈するからである。そしてこれらをうまく結びつけるよう要求された際に考え出された事例の大部分は、それをつけ加えることによって情報aを「マアちゃんはうれしい」と解釈ないですむような事例になっているように見える。つまり「マアちゃんはアイスが嫌いだった」というような事例は、情報bの解釈（以下では解釈bと記す）は肯定するが情報aの解釈（解釈aと記す）を否定するかたちになっていると考えられる。以下このタイプの事例を$a \cdot b$タイプとよぶ。この論法

---

[7] 内田伸子（1982）「Ⅲ. 文章理解と知識」佐伯胖『認知心理学講座 第３巻 推論と理解』東京大学出版会、p.162-163
[8] 内田前掲書 p.159-160,170
[9] お茶の水女子大学人間文化研究科 所属
[10]「矛盾の解消」については後に詳しく説明する。
[11] 今井むつみ・秋田喜美（2023）『言語の本質』中央公論新書 p.ⅱ,ⅲ,123-127
[12] NRI（2023）「大規模言語モデル」https://www.nri.com/jp/knowledge/glossary/lst/ta/llm 2023年7月22日閲覧
[13] 鈴木考子（1985）「一見矛盾する情報の理解過程における事例構成」『教育心理学研究』33巻2号所収 p.116-117
[14] 東京大学大学院教育学研究科 所属
[15] 久保ゆかり(1982)「幼児における矛盾する出来事のエピソードの構成による理解」『教育心理学研究』30巻3号所収 p.239-240

を押し進めて行くと、解釈aは肯定するが解釈bを否定するかたちになっているような事例があり得るはずである。久保の課題に従えば、先の表情図を「うれしくない」と解釈しないケースである。そこで再び久保のデータを見ると1例だけだが先の表情を「うれし泣き」と解したケースを見出すことができる。以下このタイプの事例を$a \cdot b$タイプとよぶ。（中略）また、理論の上では解釈aもbも肯定するがそれらの判断の同時性（aとbは同時に同じものについて判断できるとみなすこと）を否定したかたちになっている事例（$a \cdot b$タイプとよぶ）が考え出されても不思議はない。[16]

つまり「矛盾の解消」とは、最初に与えられた情報の最初の解釈を否定することで一貫性・整合性のある文章に再構築することである。そのため最初に与えられた情報文を、否定しにくいものにすると、矛盾の解消の際にその解釈を否定している事例が出現しにくくなるはずである。

鈴木の①の実験では、大学生10人に対して図1の基本例文をベースとして実験を行った。また図2のように基本例文の情報aの表現を基本例文ほど多義的に解釈できないように変えたA型例文、反対に情報bの表現を変えたB型例文、基本例文の情報a・bともにそのままで、両者の間に時間的経過を表す1文を挿入したC型例文の計4つの例文を用意した。

具体的な実験方法は次のようなものである。大学生と1対1の面接を行い、例文を1度読み聞かせ、話の中につじつまの合わない所があったか矛盾点を指摘させる。大学生には矛盾を指摘した直後にその情報解釈について「本当にそれで良いか」と念を押す。その後（イ）確かにそう言えると思うか、（ロ）そう言えるとは限らないと思うが説明できないか、（ハ）そう言えるとは限らない事を説明できる（この場合に限り大学生に説明を求める）かのいずれかひとつを選択させる。最後に例文が全部本当のことだとしたらどう理解すればよいか説明させ、矛盾を解消させる。

0 花子さんはかけっこが大好きです
1 きのうも公園で仲良しのみどりさん・のぶ子さん・たろう君・しげお君とかけっこをしました。
2 のぶこさんのお姉さんが審判になりました。
3 五人は松の下にならびました。
4 お姉さんはそこから50メートルくらい先に立って「ヨーイ、ドン」と言いました。
5 お姉さんのいる所がゴールです。
6 はな子さんは一生懸命走りました。
7 もうすぐゴールという所で, はな子さんはみどりさんに追い抜かれてしまいました。
8 はな子さんはみどりさんを追い抜くことができないままゴールインしました。
9 はな子さんはがんばったので一着になることができました。

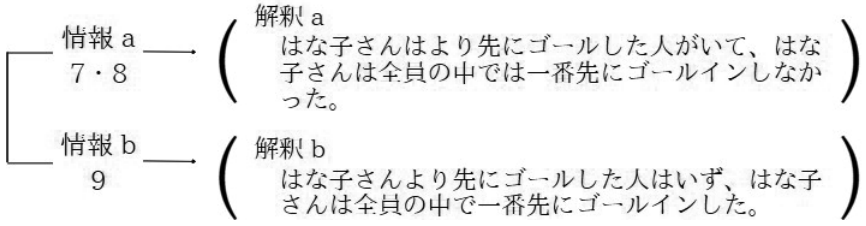

図1　矛盾の表現を含む基本例文と矛盾の構造

[16] 鈴木前掲書 p.116

A型例文
8　はな子さんはみどりさんがゴールインしたすぐ後にゴールインしました。
9　（基本例文と同じ）

B型例文
8　（基本例文と同じ）
9　はな子さんは頑張ったので、はな子さんより先にゴールインできた人はいませんでした。

C型例文
8　（基本例文と同じ）
＊　公園はもうすっかり日が暮れかかっていました。
9　（基本例文と同じ）

図2　矛盾の表現を含む変形例文

矛盾の解消の方法を説明してもらい、その回答をタイプ毎に分類し、例文の種類毎にまとめた。その結果は表1の通りである。

（単位：１人）

	基本	A型	B型	C型
$\overline{a}\cdot b$ タイプ	5	1	5	4
$a\cdot\overline{b}$ タイプ	8	9	3	4
$\overline{a\cdot b}$ タイプ	1	3	3	8

表1　例文別に見た各事例タイプ構成者数（人間）

表1の結果より、A型例文は基本例文と比べ $\overline{a}\cdot b$ タイプが減少し、B型例文は基本例文と比べ $a\cdot\overline{b}$ タイプが減少し、C型例文は $\overline{a\cdot b}$ タイプが増加した。これは実験を始める前に設定した各例文の特徴と同じである。よって、人間は最初に与えられる文の表現を変えることで、矛盾の解消の際にその文の解釈を変えることが可能であると言えそうだ。

次に、②の実験についてである。②の実験の仮説が正しければ、ヒントを与え、暗黙裏に想定している条件を意識化させることで、自分の回答に疑問を持ち、結果的に特殊な状況を考え出すことができ、解答数は増加すると仮定できる。

②の「特殊状況を考えさせる」実験では大学生32人をヒント群と無ヒント群に半数ずつ分け、図3の「1等賞の問題」を使い、1対1の面接を行った。まずヒント群のみ問題を1度読み聞かせ、ABC3つの条件（A;ルールを守っている、B;1等になれた、C;賞を受け取れた）を記入したヒントカードを提示する。その後、どちらの群も問題を1度読み聞かせ、なぜなのか答えさせる。

ⓐケンちゃんは運動会のかけっこで１番先にゴールインしました。
ⓑそれなのに、ケンちゃんは１等賞をもらえませんでした。

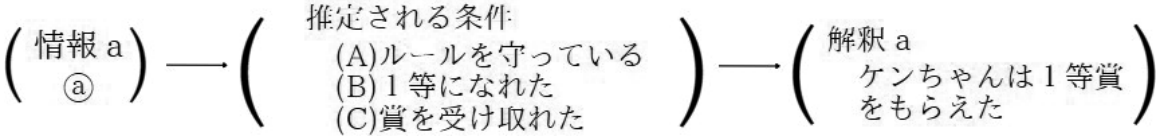

図3　一等賞の問題文と解釈の構造

大学生の回答をABCのどの「カテゴリー」に反している回答か分類し、その分類した回答を基に一人で何種類のカテゴリーに言及したか数える言及カテゴリー数と答えた数を無ヒント・ヒント群に分けてまとめた。その結果は表2の通りである。

	（単位：1） 言及カテゴリー数	（単位：1解答） 解答数
無ヒント	1.5	2.2
ヒント	2.2	3.8

表2　言及カテゴリー数と解答数（人間）

表2の結果よりヒントを与えることで言及カテゴリー数、回答数ともに増加する傾向があることが分かった。暗黙裏に想定されている条件を意識化させることで、その条件が本当に満たされているのかを疑わせ、その結果、矛盾の解消促進に繋がる。

（2）生成AIの場合

この2つの実験と同様のものを、生成AIを使って検証していく。今回の実験で使うLLMの種類はOpenAI社のGPT-3.5、GPT-4[17]、Google社のPaLM 2[18]、Meta社のLlama 2[19]の4種類である。

①の実験では鈴木の作成した4つの例文を入力し、人間に対して行ったのと同じように質問をし、出力してもらう。またPaLM 2は日本語での出力に応答しなかったため英語での検証も全LLMで行う。英語は日本語よりも開発が進んでおり、能力も高いため文章を「理解」する能力も高いと仮定できる[20]。

その結果は表3・4の通りである。

[17] OpenAI（2023）「GPT-4」https://openai.com/research/gpt-4　2023年7月17日閲覧
[18] Google（2023）「Introducing PaLM 2」
https://blog.google/technology/ai/google-palm-2-ai-large-language-model 2023年7月17日閲覧
[19] Meta（2023）「Llama 2 - Meta AI」https://ai.meta.com/llama/ 2023年7月19日閲覧
[20] OpenAI前掲ページ

（単位：1回）

		基本	A型	B型	C型
GPT 3.5	$\overline{a}\cdot b$ タイプ	3	2	1	3
	$a\cdot\overline{b}$ タイプ	2	4	2	4
	$\overline{a\cdot b}$ タイプ	0	1	0	4
GPT 4	$\overline{a}\cdot b$ タイプ	6	5	5	6
	$a\cdot\overline{b}$ タイプ	8	4	7	7
	$\overline{a\cdot b}$ タイプ	5	2	3	9
PaLM 2	$\overline{a}\cdot b$ タイプ	-	-	-	-
	$a\cdot\overline{b}$ タイプ	-	-	-	-
	$\overline{a\cdot b}$ タイプ	-	-	-	-
Llama 2	$\overline{a}\cdot b$ タイプ	3	2	3	3
	$a\cdot\overline{b}$ タイプ	4	4	5	4
	$\overline{a\cdot b}$ タイプ	0	0	1	0

表3　日本語における例文別に見た各事例タイプ構成者数（LLM）

（単位：1回）

		基本	A型	B型	C型
GPT 3.5	$\overline{a}\cdot b$ タイプ	5	4	2	3
	$a\cdot\overline{b}$ タイプ	6	4	4	2
	$\overline{a\cdot b}$ タイプ	0	0	0	1
GPT 4	$\overline{a}\cdot b$ タイプ	8	6	7	6
	$a\cdot\overline{b}$ タイプ	7	4	3	5
	$\overline{a\cdot b}$ タイプ	5	1	2	7
PaLM 2	$\overline{a}\cdot b$ タイプ	4	7	5	5
	$a\cdot\overline{b}$ タイプ	3	5	2	4
	$\overline{a\cdot b}$ タイプ	4	1	0	3
Llama 2	$\overline{a}\cdot b$ タイプ	5	4	3	2
	$a\cdot\overline{b}$ タイプ	2	3	4	2
	$\overline{a\cdot b}$ タイプ	0	0	0	1

表4　英語における例文別に見た各事例タイプ構成者数（LLM）

表3・4の結果より、人間でみられていた各例文の規則性が生成AIでは見られなかった。よってLLMに対して与える文を変えても、矛盾の解消の際にその文の解釈を変えることは難しそう

である。これは、冒頭で述べたとおり、生成AIの文章理解の方法が人間と異なるからであると考えられる。また、LLMの性能が上がるほど矛盾の解消の構成数は大きく増加し、より開発の進んでいる英語で行うと構成数はさらに増加する傾向であることも分かった。

②の実験では鈴木の作成した「1等賞の問題」を入力し、人間に対して行ったものと同様に質問をし、出力してもらう。

その結果は表5の通りである。

		（単位：１） 言及カテゴリー数	（単位：１解答） 解答数
GPT 3.5	無ヒント	2.4	4.4
	ヒント	2.8	3.8
GPT 4	無ヒント	2.1	2.8
	ヒント	2.4	3.0
PaLM 2	無ヒント	1.4	1.9
	ヒント	2.5	2.9
Llama 2	無ヒント	2.3	3.7
	ヒント	2.7	4.9

表5　言及カテゴリー数と解答数（LLM）

表5の結果よりヒントを与えることで言及カテゴリー数、回答数ともに増加しやすい傾向であることが明らかになった。よって暗黙裏に想定されている条件を意識化させ、その条件が本当に満たされているのか疑わせることで、生成AIに対しても矛盾の解消促進に繋げることができそうである。

以上2つの実験から、生成AIの文章理解の方法は人間と異なるが、矛盾の解消の構成数や言及カテゴリー数、解答数等の能力は人間と同等、もしくはそれ以上の能力を持つことが分かった。

これらの実験結果と理解の6側面を比較することで、生成AIの文章「理解」について考えられるのではないだろうか。この実験では理解の6側面のうち説明、応用、パースペクティブの3つの側面から「理解」を測っている。①の実験では矛盾している箇所を指摘し、なぜ矛盾しているのかを説明させ、その指摘の正誤を尋ね、パースペクティブを持たせて説明させる。その後文章を応用させていくことで矛盾を解消させていく。

②の実験でも文章で暗黙裏に想定されている条件を表面化し、そこから応用させることで問題の解消につなげる。このように、人間と生成AIの実験結果を比較することで、前述の3つの側面についての「理解」度合いが分かるのではないか。

1つ目の「説明する」ことに関しては、①の実験で矛盾点があるか、それはどこか、どのように違うのかを指摘させる箇所が該当する。人間は全員説明することができ、生成AIも全ての場合で指摘を行うことができていた。

2つ目の「応用する」ことに関しては、①の実験では矛盾の解消を行う箇所が該当する。②の実験では暗黙裏に想定される条件を否定させることで問題を解く箇所が該当する。人間と生成

AIのデータ数値を比較すると、どちらの実験でも生成AIは応用する能力が人間と同じ、もしくは人間よりも優れていた。

3つ目の「パースペクティブを持つ」に関しては、①の実験で最初の指摘が本当に妥当なのかと聞く箇所が該当する。人間は全員パースペクティブを持ち、（ハ）を選んでいたのに対し、生成AIはLLMの種類によって出力結果が変わっていた。GPT-3.5は日本語での出力の時、（イ）を選んでいた。そのうえ本当に妥当なのかと聞くと最初は矛盾していると言っていたのに、「指摘が間違っていた」や「矛盾している点は実際にはなかった」等、間違った主張に変えてしまう出力パターンが多く確認された。

以上のことから今回の実験では理解の6側面のうち説明・応用することは生成AIでも行うことができていたが、パースペクティブを持つことは不十分であることが分かる。なぜ生成AIはパースペクティブを持つことができないのだろうか。それは、LLMは「あなたの指摘は本当にあっていますか？」等のプロンプトに対して、異常なほど敏感だからだと考えられる。最初は矛盾していないと回答していたLLMも、あとになって矛盾していたと訂正することがあった。これは全てのLLMに該当し、パースペクティブを持たせるように促す質問をすると、むしろ逆の主張に変えてしまいやすいということだ。

Ⅳ　AIにおける「理解」概念の実現可能性

（1）理解の6側面から捉える生成AI

では理解の6側面のうち前章で触れていない解釈、共感、自己認識について生成AIの現状を整理しつつ今後の展望について言及していく。

LLMは与えられたスクリプトに応じて分かりやすい口調や会話形式等の条件をつけての説明に対応し、ほかの事象を使って変換することができる。そのため生成AIは解釈をしていると考えられやすい。しかし、解釈は「合理的ではあるが多様な解釈を見いだすために、テキストと自分の経験を行き来する。（中略）すべての解釈は、それが生まれてきた個人的、社会的、文化的、歴史的な文脈に縛られている」[21]と説明されている。つまり記号接地をせず、既有知識を持ちづらい生成AIは解釈を行っているとは考えづらい。

次に共感について考えていく。学習環境デザイナーである美馬のゆりは、AIが社会に浸透していく上で共感によって意図を理解し、そこから問題を解決していく力が重要であるとして、必要とされる人材について次のように主張している。

> 第3の時代の今、必要とされているのは、データサイエンティストです。深層学習やニューラルネットワークシステムのために必要とされています。（中略）データを収集して分析するだけでなく、データ分析に基づいて合理的かつ包括的な意思決定を行うためのシステムを開発する人です。／その次の時代に必要だと考えるのは、「共感デザイナー」（empathic designer）と私が名づけた役割の人です。ケアの倫理により課題を見出し、コンピュータで解決できるよう、データとして表現できるようにする人です。共感デザイナーとは、当事者意識を持って身の回りにある課題を見つけ、それを解決するために、AIをツールとして活用できる形で定式化し、新しいモノや仕組みを考え出し、責任を持って行動していく人のこと。その過程において、課題の置かれた状況の対立やジレンマを調整していくのです。これは、AIと共生する社会において必要な、新しい知性のあり方、「創造的共感知性（creative empathic intelli- gence）」といえるでしょう。[22]

システム開発者としてのデータサイエンティストの重要性は言うまでもないが、ここで注目したいのは「共感デザイナー」である。つまり現在の生成AIは共感という視点を持たずに出力し

---

[21] G.ウィギンズ・J.マクタイ前掲書 p.109

[22] 美馬のゆり（2021）『AIの時代を生きる―未来をデザインする創造力と共感力』岩波書店、p.125-126。下線は筆者による。

ているため、共感デザイナーという役割が必要だということだ。しかし、AIの発展によって共感デザイナーの役割をAI自身が行う可能性がある。「創造的共感知性」のようなものを定量化し、AIに実装できれば、ユーザーの意向に沿った、尋ねる前からよいと思ってもらえるような結果を出力できるようになる。自分の立場から逃れ、ユーザーの立場に立つことができれば、AIが「共感」したという状態をつくることができるということだ。

最後に自己認識について考えていく。ここで「なぜその答えを出したのか」を説明できる、能力が高いAIを「説明可能なAI（Explainable AI）」と呼ぶ。

AIは信用と信頼性がもっとも重要視されている。説明可能なAIはモデルの精度、公平性、透明性、結果の表れであり、そのAIモデル自体の信頼に繋がる。しかしAIの高度化によって、アルゴリズムがブラックボックス化[23]してきており、解釈不可能なものをいかにして人間のユーザーが理解し、潜在的なバイアスを説明できるかが求められている。

説明可能なAIという概念には、AIの出す結論が信頼に足るものなのか最終的に判断し、責任を持つのは人間であるという前提がある。しかし、AIの発展によってAI自身が「なぜその答えを出したのか」を説明し、AIモデルの説明可能性を自身が証明するようになるのではないか。それが実現可能になれば、人間が介さずとも自身を説明し、自己認識をしているかのように見せることができるということだ。

以上のことより、生成AIは理解の6側面のうちまだ2つの側面しか実行できておらず、本質的な「理解」をしているかのように見せる状態にはほど遠い。しかし、一部の側面では人間と比べて、文章理解の能力が同等もしくはそれ以上と見なしうる場合もあった。年々、人間とは異なった仕方で「理解」する、しているように見せる能力が向上しているため、数年後にはAIの「理解」が、人間のその能力を上回るようになるかもしれない。

（2）汎用AIの「理解」とは

前節までは生成AIでの話をしてきたが、生成AIはあくまでも物事の特徴を捉え、傾向を掴むことでうまく「人間っぽいこと」を出力しており、人間の入力（指示）がなければ、出力をすることもできない。

現在主流のAIは、限られた課題にのみ対応可能な特化型AIである。しかし、真のAIはこれではない。囲碁等のボードゲームをするAlphaGo Zero[24]は自動車の運転をすることはできないし、ChatGPTはソーラン節の踊り方は説明できるが、実際に踊ったり、人間の踊りに対してアドバイスしたりすることはできない。特化型AIは人間が予め想定した特定の課題しか行えない。それに対して、決められた課題だけではなく、複雑かつ未知の課題にも臨機応変に対応できる真のAIは「汎用AI」である。まるで人間のような行動をするAIは人間の立場を大きく崩すこととなるだろう。

しかし汎用AIはまだ現れていない。その要因として、「フレーム問題」と、前述した「記号接地問題」（シンボルグラウンディング問題）が挙げられる。フレーム問題は与えられた命令に対しその裏にある既有知識を「理解」することができず、記号接地問題は物事を概念としてではなく記号として認識するため、その物事の本質を「理解」できていないことを表す。

前者は、本稿Ⅲ章で生成AIに対して行った実験②でも無ヒントの値が高かったことから、生成AIが常識を持つことの難しさが分かり、後者については、Ⅱ章で触れた生成AIの文章理解の方法から同様の帰結となった。つまり、汎用AIが物事を「理解」し、人間と同じように様々なタスクをこなすためには、どれほど「理解」と見なしうるアウトプットが可能になったとしても、特化型AIの延長線上では不可能だ。

---

[23] 中身が分からないもの。今回の場合は人間がアルゴリズムや内部の仕組みが分からなくても出力できるということ。

[24] Google DeepMind（2023）「AlphaGo」https://www.deepmind.com/research/highlighted-research/alphago 2023年9月30日閲覧

[25] Carl Benedikt Frey･Michael Osborne（2013）「The Future of Employment: How susceptible are jobs to computerisation?」https://www.oxfordmartin.ox.ac.uk/downloads/academic/future-of-employment.pdf 2023年8月19日閲覧

この点については、オックスフォード大学のフレイ博士らが発表した、これから10～20年でなくなる可能性の高い職業と低い職業の一覧を見れば明確だ（表6）[25]。

表6　10～20年でなくなる可能性の高い職業と低い職業（翻訳は筆者）

	10～20年でなくなる可能性の高い職業	10～20年でなくなる可能性の低い職業
1	テレマーケター	レクリエーション療法士
2	不動産の所有・登録調査員	整備士、設置業者、修理業者の第一線監督者
3	縫製・手縫い	危機管理責任者
4	計算オペレーター	メンタルヘルス・薬物乱用ソーシャルワーカー
5	保険業者	聴覚訓練士
6	時計修理工	作業療法士
7	貨物代理店業者	歯科矯正医・義肢装具士
8	税務申告代行者	医療ソーシャルワーカー

出典）Carl Benedikt Frey･Michael Osborne（2013）「The Future of Employment: How susceptible are jobs to computerisation?」www.oxfordmartin.ox.ac.uk/downloads/academic/future-of-employment.pdf

表1から分かる通り、前者はテレマーケターやデータ収集･加工等、業務内容がパターン化しているものが多く、特化型AIによって代替可能である職業が多い。それに対し、後者は療法士や危機管理責任者、ソーシャルワーカー等、他の人間とのコンタクトが必要不可欠で、求められるタスクが広範囲に広がっている業務のため、パターン化できず、特化型AIがすぐに取って代わることは不可能である。

そこで人間の脳を模した汎用AIの実現を目指す全脳アーキテクチャ[26]が注目を集める。これは特化型 AIに対置する形で設定された技術目標であり、最初からあらゆる問題に対応できる万能なAIをつくるのではなく、人間同様に特化した知能を柔軟に習得できる能力をもつAIをつくるというものである。

言語学には、生まれたとき人間は最も基本的かつ汎用的な学習機構以外は何も持っておらず、全ての知識は外部から来る、経験に由来するという経験主義の概念が存在する[27]。それと同じようにWBAI[28]は「脳はそれぞれよく定義された機能を持つ機械学習器が一定のやり方で組み合わされる事で機能を実現しておりそれを真似て人工的に構成された機械学習器を組み合わせる事で人間並みかそれ以上の能力を持つ汎用の知能機械を構築可能である」という仮説のもと、研究･開発を行っている。知識は意味の網の目を作り、知識を支え、長い時間をかけて知識の網の目を作り上げることで「理解」できるようになる[29]と山鳥が提唱するように、知識の網の目を作れる能力をAIに組み込めればよいのだ。つまり人間が成長するのと同じように「理解」を進めることができるようになれば、AIはAIなりにその物事の意味を「理解」し、理解の6側面のそれぞれを代替的に行うことができる可能性をもつ。

Ⅴ　人間とAIの今後

生成AIは現在の能力･技術では人間と同じように文章を「理解」することはできていなかったが、「理解」する、したように見せる能力は確実に成長してきている。しかし人間と生成AIでは文章

[26] WBAI（2016）「全脳アーキテクチャとは」https://wba-initiative.org/wba/ 2023年8月11日閲覧
[27] 酒井邦嘉･辻子美保子･鶴岡慶雅･福井直樹（2022）「AIは人間の脳を超えられるか」酒井邦嘉『脳とAI―言語と思考へのアプローチ』中央公論新社、p.87
[28] 特定非営利活動法人　全脳アーキテクチャ･イニシアティブの略語
[29] 山鳥前掲書 p.182-189

理解の方法が違うため、全脳アーキテクチャ等、AIの根本から変える技術が生まれない限り総体としての人間を超えることはないだろう。言語学者の福井直樹は、人間もAIも最適化が重要な役割を持つとして、「何がどういう初期条件のもとで最適化された結果、言語能力がヒトに生じたのか、そしてその結果生じた普遍文法においては、どういった最適化原理が働いているのかを発見する」[30]ことで、人間がものを理解するとはどういうことかが分かる、としている。

今後もAIは加速度的に発展し続けるだろう。電話交換手のように新しい道具、技術の発展によって消失する職種は多数存在する。自身が「理解」する能力を超えられた人間は、現時点でも仕事を奪われることは確実だ。現在のAIは人間による入力が行われないと動かず出力しないが、自分で自我を持つかのような汎用AIが生まれた後には、真のシンギュラリティが待っているかもしれない。

（10,489字　原稿用紙26.2枚相当）

[30] 酒井前掲書 p.129-131

【参考文献および関連URL】

●新井紀子（2018）『AI vs. 教科書が読めない子どもたち』東洋経済新報社
●今井むつみ･秋田喜美（2023）『言語の本質』中央公論新書
●内田伸子（1982）「Ⅲ. 文章理解と知識」佐伯胖『認知心理学講座 第3巻 推論と理解』東京大学出版会
●久保ゆかり（1982）「幼児における矛盾する出来事のエピソードの構成による理解」『教育心理学研究』30巻3号所収
●酒井邦嘉･辻子美保子･鶴岡慶雅･福井直樹（2022）「AIは人間の脳を超えられるか」酒井邦嘉『脳とAI―言語と思考へのアプローチ』中央公論新社
●鈴木考子（1985）「一見矛盾する情報の理解過程における事例構成」『教育心理学研究』33巻2号所収
●ニュートン（2022）『ゼロからわかる人工知能』ニュートンプレス
●野村直之（2016）『人工知能が変える仕事の未来』日本経済新聞出版社
●松尾 豊（2015）『人工知能は人間を超えるか ディープラーニングの先にあるもの』KADOKAWA
●美馬のゆり（2021）『AIの時代を生きる―未来をデザインする創造力と共感力』岩波書店
●山鳥重（2002）『「わかる」とはどういうことか』筑摩書房
●G.ウィギンズ･J.マクタイ／西岡加名恵訳（2012）『理解をもたらすカリキュラム設計―「逆向き設計」の理論と方法』日本標準
●J.マッカーシー･P.J.ヘイズ･松原仁／三浦謙訳（1990）『「人工知能になぜ哲学が必要か」フレーム問題の発展と展開』哲学書房
●Carl Benedikt Frey･Michael Osborne（2013）「The Future of Employment: How susceptible are jobs to computerisation?」https://www.oxfordmartin.ox.ac.uk/downloads/academic/future-of-employment.pdf
●Google（2023）「Introducing PaLM 2」https://blog.google/technology/ai/google-palm-2-ai-large-language-model
●Google DeepMind（2023）「AlphaGo」https://www.deepmind.com/research/highlighted-research/alphago
●Meta（2023）「Llama 2 - Meta AI」https://ai.meta.com/llama/
●NRI（2023）「大規模言語モデル」https://www.nri.com/jp/knowledge/glossary/lst/ta/llm
●OpenAI（2023）「GPT-4」https://openai.com/research/gpt-4
●WBAI（2016）「全脳アーキテクチャとは」https://wba-initiative.org/wba/