ChatGPT革命

ChatGPT Revolution

国立研究開発法人産業技術総合研究所　フェロー
辻井　潤一

国立研究開発法人産業技術総合研究所　フェロー、英国マンチェスター大学教授、国際計算言語委員会（ICCL）委員長、AAMT ／ Japio 特許翻訳研究会委員長

1　前史

OpenAI が昨年 11 月に発表した ChatGPT は、その能力の高さから AI 技術の研究開発や応用の世界を大きく変革し始めている。AI は、過去にも大きな出来事によりその在り方が大きく変化してきた。ChatGPT 革命の前には、2015 年の Google 傘下の DeepMind による AlphaGo があった。

2015 年は、日本政府が AI 技術の重要性から、AI 研究センターを産総研に設立し、私がそのセンター長に就いた年である。その年の秋、AI のシンポジュムでソウルに滞在していた時、AlphaGO と韓国のプロ棋士が対戦しており、AlphaGo が勝利した。この出来事が、シンポそこのけの大きな話題となっていた。この AlphaGo は、深層学習の大きなうねりを象徴するものとなった。

実際、AlphaGO の出現の前から、私の専門とする機械翻訳の分野では、深層学習は単純な CNN（Convolutional Neural Network）から、単語列のもつ構造を取り扱うための RNN（Recursive Neural Network）、LSTM（Long Short Term Memory）といった、さまざまな深層学習技術の進化形が提案され、ニューラル翻訳の研究が、盛んになっていた。この深層学習の変革の極め付きが、ChatGPT の技術基盤であるトランスフォーマー（Transformer）であった（参考文献 1）

トランスフォーマーは、無限定的な文脈を参照するために巨大なモデルとなるが、並列処理が可能なことで、モデルの巨大さにもかかわらず、大規模な GPU クラスターを使うことで学習時間を実用可能な範囲に収めることができた。このトランスフォーマーに関する論文を Google Brain が発表したのは 2017 年であり、その後、この計算枠組みは、わずか 5 年の間に様々な分野で活用されることになった（参考文献 2）。

例えば、Google 傘下の DeepMind が 2018 年、2020 年に発表した AlphaFold1、AlphaFold 2 は、タンパク質構造予測の精度を画期的に改善し、生命科学の研究分野を革命的に変化させた。この画期的な性能を示した AlphaFold2 は、トランスフォーマー技術に基づく（参考文献 3）。

2　商業的な価値と競争

ChatGPT を開発した OpenAI 社は、2015 年に設立される。彼らは、AGI（Artificial General Intelligence）という、特定分野の知能ではなく、人間知能と同じように幅広い分野で知的な能力を示す汎用知能の実現を目指す、としていた。同社の設立には、テスラの E. マスク氏も大きな役割を果たす。

ただ、マスク氏は、2018 年に OpenAI 社から手を引く。理由は定かではないが、Google 傘下の DeepMind や Google Brain[1] が前述のような画期的な成果を上げているのに対して、OpenAI が沈滞しているとの印象を持ったため、という人もいる。その

1　この 2 つのグループは、本年 3 月に Google DeepMind に統合された

後、OpenAI は、マイクロソフトなどの援助を受け、ChatGPT という、画期的なシステムを発表する。

ChatGPT は、Web 検索、チャットボットの世界を大きく変えるだけでなく、言語での会話を通してプログラム作成を行うなど、サイバー空間での AI 技術を大きく変革している。Web 検索、チャットボット、あるいは AI 技術一般が持つ膨大な市場性から、ChatGPT が出現した直後から、Google や Amazon、Apple、Meta（Facebook）、中国の IT 企業らが一斉に、その基盤となった大規模言語モデル（以下、LLM：Large Language Model、GPT は LLM の典型的なもの）の開発に乗り出している。マスク氏も、より信頼性の高い LLM の開発に投資を行うとアナウンスするなど、過去の経緯から見ると皮肉な状況となっている。

昨今、ChatGPT 的な技術の大きな可能性とともに、懸念も表明されている。技術的には、

（1）数百億から千億個を超えるパラメータを持つ LLM の内部動作が不透明なブラックボックスになっていること、
（2）膨大なパラメータを調整する学習ために、大規模な言語データや大きな計算リソースを必要とすること
（3）ハルシネーション（hallucination―幻覚現象）、誤情報など、信頼性に欠けること

がある。

この技術の巨大な市場性から、LLM 構築を行う企業が技術の囲い込みを目指す傾向もあり、技術のブラックボックス性が顕著になっている。また、膨大な数のパラメータを最適に調整するためには、大きな学習データ（テキスト）が必要になる。後述するように（3 節参照）、LLM の基本機能は、特定なテキストが与えられたとき、それに続く自然な後続テキストを予測すること、である。「自然な」後続テキストの予測には、実際に人間が使っている言語テキストを使う必要がある。すでに世に存在するテキストで学習するので、テキストに含まれる個人情報やテキストの著作権の問題が生じる。しかも、ChatGPT のように自らが新たなテキストを作り出す生成型 AI（Generative AI）の場合には、学習テキストに類似したテキスト（あるいは、部分的には同じテキスト）を自らが作りだしてしまうことから、個人情報・著作権の問題はより深刻になる。

3 LLM とトランスフォーマー技術、生成 AI

言語モデルとは、どういうものかをすこし説明しておこう。

図 1 は、言語モデルの簡単な例である。「英国の首都は」という先行文脈があると、「パリである」や「東京である」に比べて「ロンドンである」の確率が高くなる。先行文脈から後続するテキストが予測できる。先行文脈を長くとると、その予測の精度が上がるが、長い先行文脈がどのように後続テキストの生成に影響するかを学習するためには、膨大な学習データが必要となる。

確率言語モデルは、語の並び方の自然さを確率でとらえる。例えば、（apples,eat,people）という単語組があると、<people・eat・apples> が <people・apples・eat> とか <apples・people・eat> とかの並びよりも確率が高くなる。英語では、名詞・動詞・名詞の並びが名詞・名詞・動詞よりも自然という文法規則が、確率的言語モデルに反映される。また、<apple・

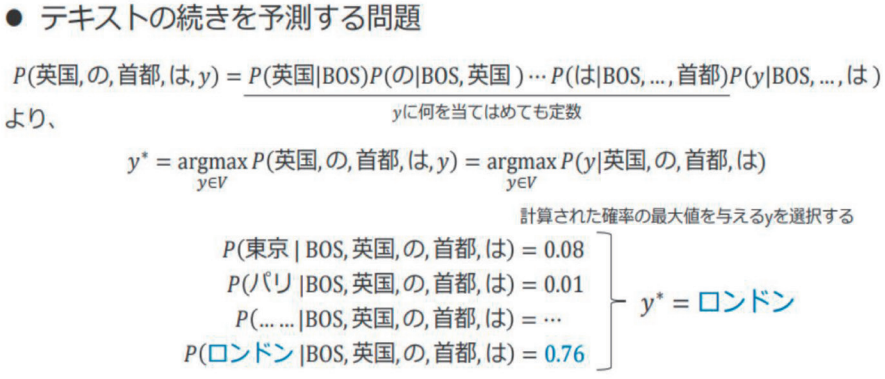


図 1　確率言語モデル（岡崎直観教授：東工大提供）

eat・people> よりも <people・eat・apple> のほうが確率が高くなるだろうから、「人間が果物を食べる」ほうが「果物が人間を食べる」よりも自然であるという、いわば、世界に関する知識も言語モデルに反映される。

このような確率言語モデルが、機械翻訳での自然な訳文を出力するのに有効なことは、想像できるだろう。

日本語文「人々はリンゴを食べる」を英語に訳する場合、(人々:people)・(人々:boys) …(リンゴ:apple)..、(食べる:eat)(食べる:have).. といった単語対応が持つ確率と英語での単語並び <people・eat・apples>、<eat・people・apples>、<boys・apples・eat> の確率の積が最大になるものを選択すればよい[2]。文法や知識といったものを規則として明示的に取り扱う必要はない。統計的機械翻訳は、これらの確率を大規模なテキスト集合から計算する。これに対して、ニューラル翻訳は、この確率に相当するものを計算するニューラルネットワークを学習によって構築する。

翻訳では、翻訳の相手言語(出力文)を生成するときに、それまでに生成した先行文脈を使って次に続く自然な表現を選択する。翻訳は、それだけではなく、入力文が出力の生成過程を制御している。たとえば、<people・eat> の後に orange や fish ではなく、apple を選択するのは、入力文 < 人は・リンゴを・食べる > の「リンゴ」があるから、である。この部分が「魚」であれば、後続は fish となろう。このように、翻訳文の生成では、それまでに出力された先行文脈だけでなく、入力文の特定の部分(この例では、「リンゴ」)が後続表現の選択に影響をもつ。この入力文のどの部分を見るかを示す機構は、注視機構(Attention)と呼ばれる。

当初の RNN や LSTM を使ったニューラル翻訳では、この注視機構は、後続表現を出力する際に入力文のどの部分に注目するかを特定するために使われていた。現在の LLM は、この注視機構をそれまでに出力されている先行文脈から後続表現を選択するのに適用する。それまでに出力された先行文脈は後続表現の選択に影響を与えるが、その影響は必ずも直前の文脈だけではない。後続する言語表現の生成に、表面上は任意に離れた表現の影響をとらえる機構として、注視機構を使う。自らが生成した先行文脈への注視機構という意味で、自己注視機構(Self-Attention)という。先行する単語並びが持つ言語構造を計算機構に取り込もうとした RNN や LTSM に代わって、単語の並び方をいわば無視して、すべてを注視機構で統一的に取り扱う(参考文献1)。

トランスフォーマーは、この自己注視機構により、単一言語の言語モデルにおいても、無限定的に長い文脈から後続する自然な言語表現を生成する強力な言語モデルとなった。たとえば、日本語での長い対話の系列から、それに後続する自然な対話系列を生成することも同じ言語モデルで取り扱える。

また、翻訳における入力文は、LLM が生成する出力を制御している、とみることができる。入力文中の <apple> は出力中の < リンゴ > を出すこと、また、<people・eat・apples> という入力文は、< 人々は・リンゴを・食べる > を出力せよ、という指令をトランスフォーマーに与えている。入力を LLM への指令とみる立場は、現在、プロンプト・エンジニアリングとして積極的に使われている(7節)。

4 トランスフォーマーの訓練

もちろん、この小文でトランスフォーマー技術の詳細は述べられない。たとえば、前節では、単語そのものが取り扱われるように説明したが、実際は、単語の特徴をとらえた分散表現(ベクトル)が基本となる。また、トランスフォーマーの計算が高次元の行列演算となり、この演算が GPU での高速並列処理で可能なことなど、重要な技術的革新は割愛した。これらの特徴から、トランスフォーマーは、

(1)無限定的な文脈を取り扱いながら、実現可能な計算時間で実現できる技術
(2)単語を基本とする言語モデルだけでなく、画像や音声の処理といった汎用の計算枠組み

となった。

トランスフォーマー技術は、翻訳の場合、入力(翻訳の場合は、入力文)を出力(翻訳の場合は、翻訳文)に変換する変換器であった。対話の系列(入力)から、それに後続する対話系列(出力)を生成するのも、変換

2 簡略して説明しているが、実際には原文と翻訳文との間での並びとしての対応の仕方(アライメント)の確率も掛け算される

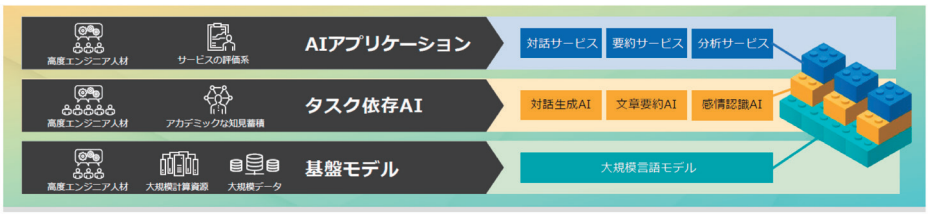


図２　ChatGPT の３層構造

器である。前節のプロンプトで述べたように、現在の ChatGPT では、入力はトランスフォーマーへの指令の役割も担っているが、入力を出力に変換するという基本の枠組みは変わっていない。

ChatGPT は、この変換器で作られた LLM（すなわち、GPT）を基盤にする。ただ、大規模な LLM があれば、現在我々が目にする ChatGPT が即座に実現できるわけではない。

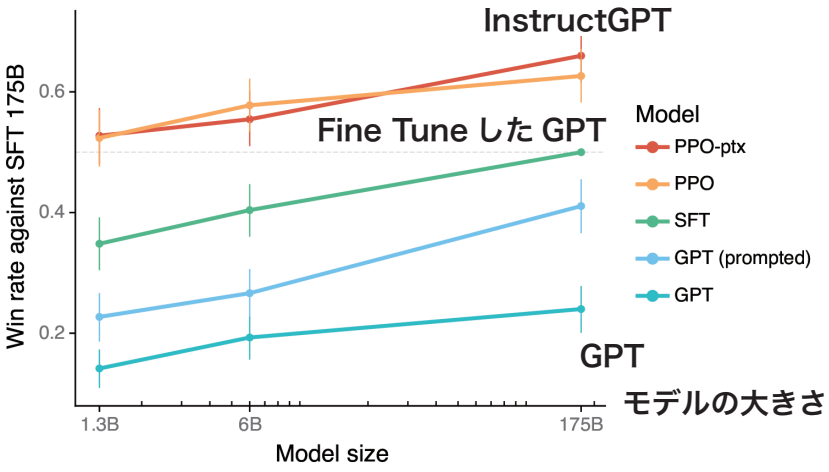
図２に、ChatGPT の３層の構成を示す。GPT は、入力を先行文脈としてそれに後続する自然な表現を生成する基本の機能を持つ。機械翻訳では、入力文と出力文の対を大量に用意することで、変換器（トランスフォーマー）を翻訳というタスク用に訓練する必要がある。同じように、たとえば、長いテキストからその自然な要約をするタスクを実現するには、そのタスク用に LLM を訓練する必要がある。

GPT（Generative Pre-Trained Transformer）の Pre-Trained は、事前学習済みの意味である。GPT は、自然な表現を作り出す能力をあらかじめ訓練された事前学習済み言語モデルではあるが、翻訳や要約、対話のような特別なタスク用にこの能力を使うためには、そのための訓練する必要がある。

事前学習モデルは、多様なタスクに対応できる基礎的な能力は持つが、その能力を特定のタスクに使うには訓練が必要である。この訓練を行う層が、図中の第２層である。第３層はトランスフォーマーへの変換の指示を入力に入れ込む層（プロンプト・エンジニアリングの層）である（７節）。

第２層の訓練は、機械翻訳の場合には、正解例（すなわち、人間の翻訳家が作った翻訳対の集合）を学習データとして与える。特定タスクの正解例を与えて訓練することは、調整（Fine Tuning、略して FT）と呼ばれる。特定のタスク用の調整には、そのタスク用の正解データを用意する必要がある。翻訳では、人間の翻訳家が作った翻訳が大量にあるので、これを使うことができる。

これに対して、長いテキストの抄録、対話での応答などのタスクでは、正解データがあまりない。また、そもそも、抄録や対話の応答文として何がよい抄録や応



横軸：言語モデルのパラメータ数、右ほど大きい
縦軸：モデルの出力を調整済みのモデル(175Bパラメータ）と比較した時の優劣
調整（Fine Tuning）やRLHFを行わないGPTは、サイズが最大のものでも、調整した最小のGPTよりもユーザ満足度は低い
RLHFまで行ったモデル（InstructGPT)は、最小のもの（1.3Bパラーメータ）でも、なにも訓練されていない最大規模のGPTよりもはるかに優れた性能

図３　訓練の効果

答文なのかは明確でない。OpenAI では、システムが作り出す抄録を人間に比較させて判断させ、その判断に合うように強化学習で LLM を調整している。OpenAI は、これを「人間（Human）からの Feedback を使った強化学習（Reinforcement Learning)」（略して、RLHF）と呼び、正解データの作成コストを削減し、良い成果が上げられたと報告している。

FT や RLHF による LLM の訓練過程を、「LLM の挙動をそのタスクに対して人間が持つ価値観に合わせる」という意味で、すこし大げさな言い方で、「LLM と人間の価値観を合わせる過程」（alignment 過程）と呼ぶ。実際、この Alignment の効果は大きく、Alingment した比較的小さな LLM（1.3Billion パラメータ）の方が Aligment していない大規模 LLM（175Billion）よりもユーザ満足度が高かったとの報告がある（参考文献 4）（図 3)。

5 ブラックボックス、コスト、バイアス

ChatGPT の最下層 GPT と呼ばれる LLM は、GPT3 で 1750 億の内部パラメータを持つ。この大量のパラメータが GPT 3の動作を規定するが、それらを人間が解釈することは不可能である。GPT3 は、人間（設計者）がその内部の動きを理解できず、巨大なブラックボックスとなる。

この大量のパラメータの学習には、FT や RLHF のような人手による学習データの構築は必要ではない。世に存在する膨大なテキスト自体が、人間が作った正解として存在し学習データとなる。その正解が生成されるようにパラメータを調整すればよい。データそのものが教師になり、正解データを必要としない自己教師学習（Self-Supervised Learning）が可能である。

ただ、この学習には、膨大なテキストが必要となる。この膨大なテキスト集合が、内部の機構はわからないが、LLM の外的挙動を決めるために、世にあるテキストをただ集めるだけではすまない。不適切な内容を含むテキスト、重複の大きなテキストなどを取り除く必要がある。これをデータ洗浄（Data Cleansing）という。膨大なテキストの収集、それを使った LLM の学習という大きな計算コストがかかる上に、このデータ洗浄にも大きなコストがかかる。

前節で述べた FT や RLHF による Alignment 過程も、大きなコストがかかる。LLM の自己教師学習に比べると、人的なコストが大きく、OpenAI の Alinment チームには、LLM の構築チームとほぼ同じ規模となっている。

「AI システムの価値を人間の価値に合わせる」といっても、人間の価値も一色ではなく、さまざまである。例えば、RLHF での出力の優劣の判断も、政治的･人種的･ジェンダー的なバイアスがかからないように、判断する人間を訓練したり、不適切な作業者は除いたり、する必要がある。

OpenAI では、ChatGPT の実ユーザの挙動をモニターし、要求されるタスクに応じて FT、RLHF による訓練を行っている。この部分に大きな力を注ぐことで、ChatGPT の優れたパーフォーマンスが達成されているようである。実際、他所で構築された同規模の LLM と比べると、例えば、長いテキストの抄録作成などのタスクで、ChatGPT は明らかにより優れた性能を示している。実ユーザの挙動に合わせた第 2 層の訓練、第 3 層でのプロンプティングの作成が、ChatGPT のパーフォーマンスを支えている。

データ洗浄、LLM の学習、訓練層での学習データなど、いずれも、大きな計算コストと人的なコストがかかる。このこととから、現在のところ、ChatGPT に追従した技術開発は、巨大 IT 企業にのみ可能なものとなっている。さらに、技術の市場価値の大きさやテキストデータの著作権・個人情報の問題から、これらの巨大企業によるデータや技術の囲い込みがおこっている。これが、巨大な LLM のブラックボックス性をさらに悪化させている。

現在、多くの公的機関、一部の企業がこのブラックボックス性に対抗して、オープン性・透明性を強調した技術開発を行いつつある。技術の負の側面を解消する技術研究には、オープンな環境の構築が不可欠であり、その展開が期待される。

6 ハルシネーション、誤情報 ― 知識の貯蔵庫か、事前学習モデルか

現在の ChatGPT の欠点、懸念事項として、その出力が誤情報を含み信頼できない、ことが挙げられる。典型的な現象に、「事実ではないことを自然なテキストで

出力する」ハルシネーション（hallucination―幻覚現象）がある。たとえば、ChatGPT に特定の個人の経歴情報を要求すると、きわめて自然なテキストではあるが、所属する機関を間違ったり、その人の業績ではないことを含んだりするテキストが出力される。現存しない人物の情報を、あたかも実在の人物のように説明するテキストが出力されることもある。

幻覚現象や誤情報の問題は、大規模言語モデル（LLM）が、

（1）自然なテキストを生成する言語モデル
（2）膨大な知識の貯蔵庫

という二面性を持つことから生じている。

自然なテキストを生成するモデルは、<people・eat・apples> が <apples・eat・people> より自然であるとして、生成する。言語モデルは、能動的な活動主体 <people> が <eat> の行為者になるのは、そうではない静的な物 <apples> が行為者になるよりも自然であるという、意味的な傾向をとらえている。

同じように、「X 氏が Y 社の社長である」は、「ある人間がある組織の長」という意味的に自然な表現にであり、言語モデルとして、これを生成することに問題はない。ただ、意味的な自然さと、事実であることは同じではない。そういう事実がない場合には、「知識の貯蔵庫」としての出力としては妥当ではない。誤情報となる。

LLM を言語モデルとしてみる立場は、LLM は膨大なテキストから言語としての自然さを学習したものであって、膨大な「知識の貯蔵庫」ではない、と考える。Web 中には真理性が疑わしいテキストも氾濫している。ただ、真理性がないテキストであっても言語としての自然さを学習するには役に立つ。ただ、それを使って学習された LLM を「知識の貯蔵庫」として使うことはできない。

膨大なテキスト集合から構築された LLM を知識の貯蔵庫として使いたいという応用が、多くある。Web 検索の代わりに ChatGPT を使うのがその典型である。実際、意味的な自然さと知識としての正しさの差は微妙であり、GPT の出力を知識の取り出しとみなせる場合も多いが、この方向性の差が誤情報や幻覚現象につながる。

真理性が判断しずらい誤情報の危険性は常にある。LLM が生成する「薬剤 X は症状 Y を示す病疾患 Z に有効である」といったテキストは自然ではあるが、真理性は保証されない。LLM の出力は、

「たとえ自然であっても、専門家が持つ明示的な知識や真理性が吟味された知識に基づいたものではないこと、に注意する必要がある」。

二面性を解消する方法の一つは、自然さを受け持つ LLM と知識の貯蔵庫の役割を分離することであろう。我々は、ヨーロッパと日本の共同プロジェクト（e-Vita）で、会話を通じて高齢者に健康に関する指導を行うシステムを開発している（図４）。健康指導の会話で、誤情報を与えてはならない（参考文献５）。

このプロジェクトでは、健康指導に関する「知識の貯蔵庫」として、WHO・NIH・ハーバード大学医学部

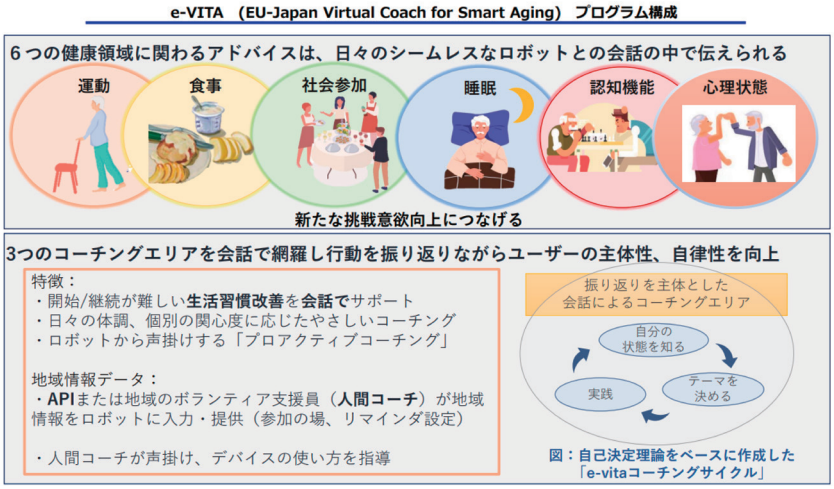


図４　e-Vita プロジェクト

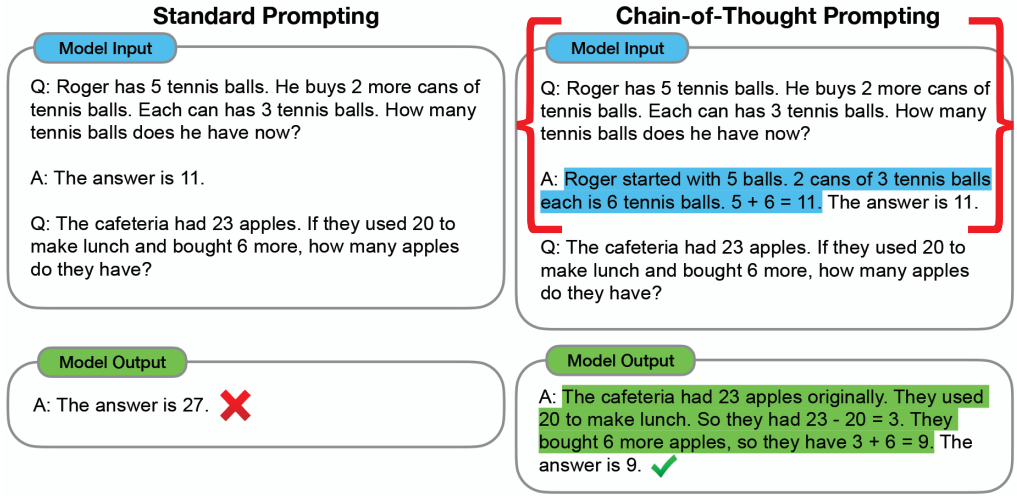


図５　CoT プロンプトの例（左側）

などが出しているガイドラインの集合を使い、この「知識の貯蔵庫」からの情報だけを使って、自然な言語で応答する。自然な応答や長い説明を簡略化する機能を ChatGPT に行わせ、望ましい結果を上げている。

「知識の貯蔵庫」と言語モデルとを分離し結合することは、LangChain（参考文献６）というオープンソースのシステムで容易に実現できる。知識の貯蔵庫と言語モデルの役割分担は、知識の貯蔵庫としては社内文書の集合や通常の Web 検索[3]の結果を使い、その内容の抄録や取り出しに LLM を使うといった形で、今後、さまざまな分野で活発化していこう。

7　LLM、プロンプト、トランスフォーマーの可能性

初期のニューラル翻訳で使われた RNN や LSTM は、言語が単語の並びであり、先行する限定された範囲の単語並びを文脈としてテキストを生成する計算機構であった（3 節）。これに対して、現在のトランスフォーマーは、

（1）単語の並びが持つ順序を重視せず、無限定的な範囲を文脈として自然なテキストを生成する、
（2）入力にトランスフォーマーへの指示の役割（プロンプト）を持たせる

という変革を行った。

この変革によって、トランスフォーマーは、言語間の単語列を変換する変換器から、言語テキストに限らないデータ間にみられる関係をとらえる汎用枠組みとなり、多様な変換作業が可能になった。例えば、仕様記述のテキストとプログラムといった異種の情報空間の相互関係をとらえることもでき、プログラム開発の補助ツールとして ChatGPT の応用が広がっている。

言語モデルとしての LLM は、言語としての自然なテキストを出力する。これに対して、トランスフォーマーは、この LLM の動作を入力に含まれる指令に従って入力から出力への変換を行う。

翻訳では、注視機構によって入力文の関連個所に注目し、それまでに生成された翻訳に後続する単語列を生成していた。すでに述べたように、トランスフォーマーは、入力テキストの単語の列という性質を捨象して、入力の関連個所を注視することで、出力の生成を制御する。

言い換えると、入力は、翻訳の場合のように変換される内容そのものである必要はなく、トランスフォーマーに行わせたい仕事の指令を含めることができる。例えば、「次の英語を日本語に翻訳してください。People eat Apple」とすると、LLM は、入力中の「次の英語を・・翻訳してください」から、変換器としての動作を変える。同じように、「次の長いテキストを短くしてください。」を前に付け加えることで、トランスフォーマーは、長いテキストを短いテキストに変換する。

この「入力中の言語で表されたユーザの意思を理解したかのように出力を生成する」機能がトランス

3　Web 検索を「知識の貯蔵庫」として使用しても誤情報の問題は避けられないが、信頼できるサイトかどうかのフィルターや変換器の出力に情報の源となったサイトを付加することで、信頼度は向上できる

フォーマーに、汎用人工知能（Arificial General Intelligence）の印象を持たせている。言語は、特定の分野に依存しない汎用性の高いメディアである。この汎用メディアで表現された仕事の指令を理解して、汎用メディアで表現された情報を操作する汎用知能が実現された、という印象を与える。

さらに、入力の一部に実行させたい仕事に関する情報を入れ込む手法（プロンプティング）は、Chain-of-Thought（CoT）と呼ばれる手法でさらに強力になっている。例えば、算術の応用問題を解かせる場合、類似の問題を解く具体的な手順をプロンプトに含めて与えると、その手順を汎化して与えられた問題を解く、ことが報告されている（参考文献 7、図５）。

複雑な問題をとく手順（CoT）の例をプロンプトに含めることで、その手順を汎化して、類似の問題を解く汎化能力は、千億を超えるパラメータを持つ大規模 LLM で初めて現れる能力だとの報告がある。LLM が巨大化すると、なぜ、このような汎化能力が発現（Emerge）するのかは、まだ解明されていない。巨大 LLM がこのような汎化の能力を発現させることは、巨大な LLM の大きな可能性を示唆するとともに、今後、解明すべき謎、理論的に解明すべき謎の典型的となっている。

8 今後に向けて

本稿では、ChatGPT 革命とはなにか、を駆け足で見てきた。この革命が、次の AI 技術だけでなく、人々の仕事や生活など未来社会の在り方に大きな影響を与えることは確かであろう（参考文献 8）。そういった議論は本稿の範囲を超える。ここでは、個人的な感想として、次の２点を挙げておこう。

（1） 超知能

「ChatGPT は、OpenAI が目指す汎用 AI（AGI）、人間知能を超えた知能の出現、R.Kurzweil らのシンギュラリティへの第一歩」とする論調があるが、個人的には、これらはハイプ（Hype）だと思う。

言語による指令をうけ、言語によって表現されたテキストを操作して、指令に従った出力をだす。あたかも、AI が言語での指令を理解し、システム中のテキストを理解して出力を作りだしているように見える。この振る舞いは、特定分野に依存しない知能を持ったかのようである。ただ、本稿で述べたように、言語による入力中の指令に応じてどのように出力するかは、FT や RLHF という大量のデータにより訓練された結果である。

また、そもそも、人間が作り出したテキストから学習した LLM は、幻覚現象・誤情報にみられるように、テキスト中の情報の真理性の吟味やそれらに対する価値判断を行う能力、理解の能力は持っていない。言語によって記述された対象を吟味・分析し、自らの価値判断に基づいて新たな知識を作り出す能力はない。

人間の対象理解の結果を言語化した、いわば、対象理解の結果としの影の世界を、自覚的な情報の取捨選択なしに、操作しているに過ぎない。

ChatGPT 的な技術が社会にもたらす負の部分は、道具としてそれを使う人間側の問題であり、負の部分を解決する社会的、技術的な方策を考えていくべきであろう。超知能、汎用 AI、シンギュラリティといったハイプに逃げ込んではいけない。

（2） 実世界と知識の世界

言語の世界は、実世界と知識の世界の中間にあって、この２つの世界を反映した影の世界である。ChatGPT は、この影の世界を対象にした AI 技術である。今後は、その外にある２つの世界と言語 AI をどのようにつなげていくかが課題となる。画像や音声、音響信号など、実世界からのセンシング情報に基づく AI、あるいは、ロボット・自動走行車のように実世界に働きかける AI と言語 AI を結び付けていくことは、直近の技術課題となる。マルチモーダル AI、知的ロボットとの連携である。

また、人間は自らを取り巻く実世界、社会、生活環境という対象を理解するための科学やそれを操作する技術の体系を構築してきた。対象に関する合理的な理解の世界をつくってきた。真理性の保証された法則性、規範に関する明示的な知識を言語 AI と結びつけることは、ハルシネーション・誤情報といった負の側面を解消するための大きな研究課題となろう。

参考文献

[1] A.Vaswani, N.Shazeer, et al. : Attention is all you need, 31st Conference on Neural Information Processing Systems（NIPS

2017）, Long Beach, CA, USA.2017
[2] R. E. Turner : An introduction to Transformers, arXiv:2304.10557v3, July, 2023
[3] A.W. Senior et al.: Improved protein structure prediction using potentials from deep learning, Nature 577 706-710 doi:10.1038/s41586-019-1923-7, 2020
[4] J.Leike, R. Lowe, et al.: Training language models to follow instructions with human feedback, OpenAI, 2022（https://cdn.openai.com/papers/Training_language_models_to_follow_instructions_with_human_feedback.pdf）
[5] e-Vita Website: https://www.e-vita.coach/
[6] LangChain Website: https://www.langchain.com/
[7] J.Wei,X.Wang,et al.: Chain-of-Thought Prompting Elicits Reasoning in Large Language Models, arXiv:2201.11903v6（cs.CL）, 2023
[8] L.Weidinger, J.Meller, et al.: Ethical and Social risks of harm from Language Models, DeepMind, arXiv.2112.04359, 2021