SE Report

テキスト生成 AI の仕組みとその利活用に関わる話題

横浜国立大学大学院環境情報研究院　教授
森　辰則 Tatsunori Mori

1. はじめに

2022年11月に OpenAI 社が ChatGPT を公開すると、瞬く間に利用者が急増し、世の中にテキスト生成 AI の利用が広がっていった。そして、テキスト生成 AI は、我々の日常生活のみならず、各界の DX において重要な役割を担うようになった。また、AI 関連企業の隆盛は、株価の上昇等、経済にも大きな影響を与えている。このように、我々の日常に深く浸透しつつあるテキスト生成 AI であるが、その仕組みを理解して利用している利用者はそれほど多くはないのではなかろうか。本稿では、テキスト生成 AI の仕組みを平易に解説しつつ、その利活用に関わるいくつかの話題を提供する。

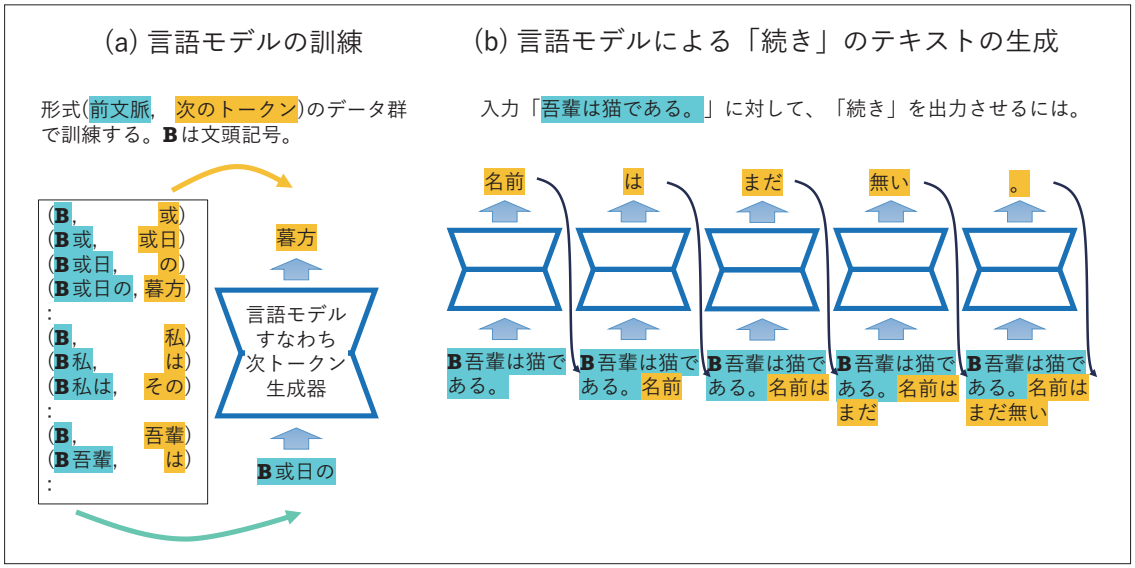
2. テキスト生成 AI の仕組み

テキストを対象とした生成 AI として広く知られているものは、ChatGPT であろう。ChatGPT という名前のうち、“Chat” はそのまま「おしゃべり、会話」を意味し、利用者と対話型のやり取りができることを示している。それでは、“GPT” とは何であろうか。GPT は、“Generative Pre-trained Transformer” の頭文字をとったものである[1)]。「生成的な事前学習」をした “Transformer” という意味になる。

まず、「生成的な事前学習」について述べる。ChatGPT のようなテキスト生成 AI は「大規模言語モデル（Large Language Model：LLM）」とも呼ばれることがある。言語モデル、正確には確率的言語モデルとは、ある語句が w_i で表される場合、与えられたテキスト $w_1w_2w_3\cdots w_{n-1}$ に対し、次に現れる語句 w_n をその出現確率とともに計算する仕組みである[2)]。これは、日本語や英語などの言語の性質に由来し、特定の仕事（タスク）に依存しないため、大量のテキストを使用して、**図1（a）**のようにあらかじめ仕込むこと、すなわち、事前学習をすることができる。これが、大規模言語モデルであり、このような事前学習を「生成的な事前学習」と表現している。

つぎに、“Transformer” について説明する。Transformer は Google の研究者たちが提案した、確率的言語モデルを実装する一手法である[3)]。複雑なニューラルネットワークにより実現されているが、事前学習により一度確率的言語モデルが学習されてしまえば、それを使った処理は単純である。入力テキスト $x_1\cdots x_n$ が与えられたら、次に続く語句 y_1 を確率的に推定し、続いて、新たに得られた語句 y_1 を入力テキストの後に追加して $x_1\cdots x_n y_1$ として、その続きの語句 y_2 を同様に推定する。この過程を繰り返すことにより、**図1（b）**のように入力テキスト $x_1\cdots x_n$ に対して、テキスト $y_1\cdots y_m$ が生成されていく。これは、頭の中にある情報に基づいて、入力テキストの続きを創造する「小説家」のような存在であると考えると分かりやすい。

図 1　言語モデルの訓練と利用



3. テキスト生成 AI への指示としてのプロンプト

一見するとただの言語データであった大規模言語モデルから、今日のテキスト生成 AI へ至る過程において、優れた着眼点があった。それは、大規模言語モデルが自由に文章を生成する「小説家」としてだけではなく、様々な仕事（タスク）をこなす「秘書」としても活用できることを見出したことである[4]。つまり、望むタスクの内容と作業対象を入力テキストとして与え、続きのテキストを生成させることで、その中に作業結果を得ることができるというものである。テキスト生成 AI として言語モデルを用いる際には、利用者からの指示に対して適切な回答が出力されるように、事前学習の際に一般的なテキストのほかに、人間からの作業指示とそれに対する出力の対も与えられる。これは、Instruction Tuning と呼ばれる。さらに、言語モデルの学習対象はテキストであれば何でもよいので、自然言語のテキストに限る必要はない。例えば、たくさんのコンピュータプログラム（ソースコード）を学習させることで、利用者の要望に応じたソースコードを生成可能である。

利用者が望むタスクの内容と作業対象を含めた一連の入力はプロンプト（Prompt）と呼ばれる。テキスト生成 AI の利用においては、利用者が望む作業や作業対象をどのような形でプロンプトに仕立てるかが非常に重要である。そのため、プロンプトエンジニアリングという、プロンプトの在り方を議論する学問分野が形成されつつある。プロンプトエンジニアリングに関するガイド類を**表1**に示す。入門書として、**表1**の DAIR.AI によるガイドを一読することをお薦めする。また、OpenAI 社も、ChatGPT を利用する際に適切な回答を得るためのベストプラクティスやガイドを提供している。

表 1　プロンプトエンジニアリングに関するガイド類

提供者	表題	URL
DAIR.AI	Prompt Engineering Guide	https://www.promptingguide.ai/jp
OpenAI	Best practices for prompt engineering with the OpenAI API	https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-the-openai-api
OpenAI	Prompt Engineering	https://platform.openai.com/docs/guides/prompt-engineering

4. ChatGPT を用いたテキストの構造化事例

ここでは、すこし複雑なプロンプトの例として、テキスト情報の構造化を紹介する。共同研究などを通じてうかがうところによると、企業では文書情報の有効活用が急務となっている。過去事例がテキストとして多量に蓄積されているものの、その大半がプレインテキストであるために通常のテキスト検索による利用にとどまっているためである。それらのプレインテキストに対して、種類を表す分類タグを付与したり、登場するモノ・コトを種類毎に抽出したり、事象の原因や結果といった重要事象を抜き出したりすること等、いわゆる、文書の構造化を行えば、利活用が進むことが期待される。大規模言語モデルを用いて適切なプロンプトを与えればある程度の精度でこのような文書の構造化が行える。例として、失敗知識データベース[5] に掲載されている、ある事例概要を対象として構造化を行ってみる。**図2（a）**が入力したプロンプトである。ここでは、参与者（Participants）、機器（Machines）、他の対象物（Objects）、場所（Places）、個別の話題（Specific topics）を表す表現をテキストから抽出するとともに、文書の分類ラベルとして使える一般的な主題（General themes）を推定するように指示している。さらに、事故原因、ならびに、人的被害に関する重要文を抽出させている。**図2（b）**は、ChatGPT 4o の出力である。この事例では、良好な処理がなされている。

5. 様々なテキスト生成 AI

Transformer が Google の開発によるものであることを踏まえれば、Google が GPT と同様のテキスト生成 AI を開発しているのではと考えることは自然であろう。実際、世の中には数多くの大規模言語モデルの実装が存在しており、それらのほとんどが Transformer を基盤としている。これらはいずれも大量のテキストによって事前学習されている。主な違いは、Transformer のどの部分を利用しているか（二つの主要な部分がある）、複雑さの指標であるパラメーター数、学習に用いたテキストの種類と量、Web 検索など外部システムとの連携の有無などにある。ここでは、ChatGPT 以外で一般の利用者向けに対話型の仕組みを備えたものをいくつか**表2**に挙げる。

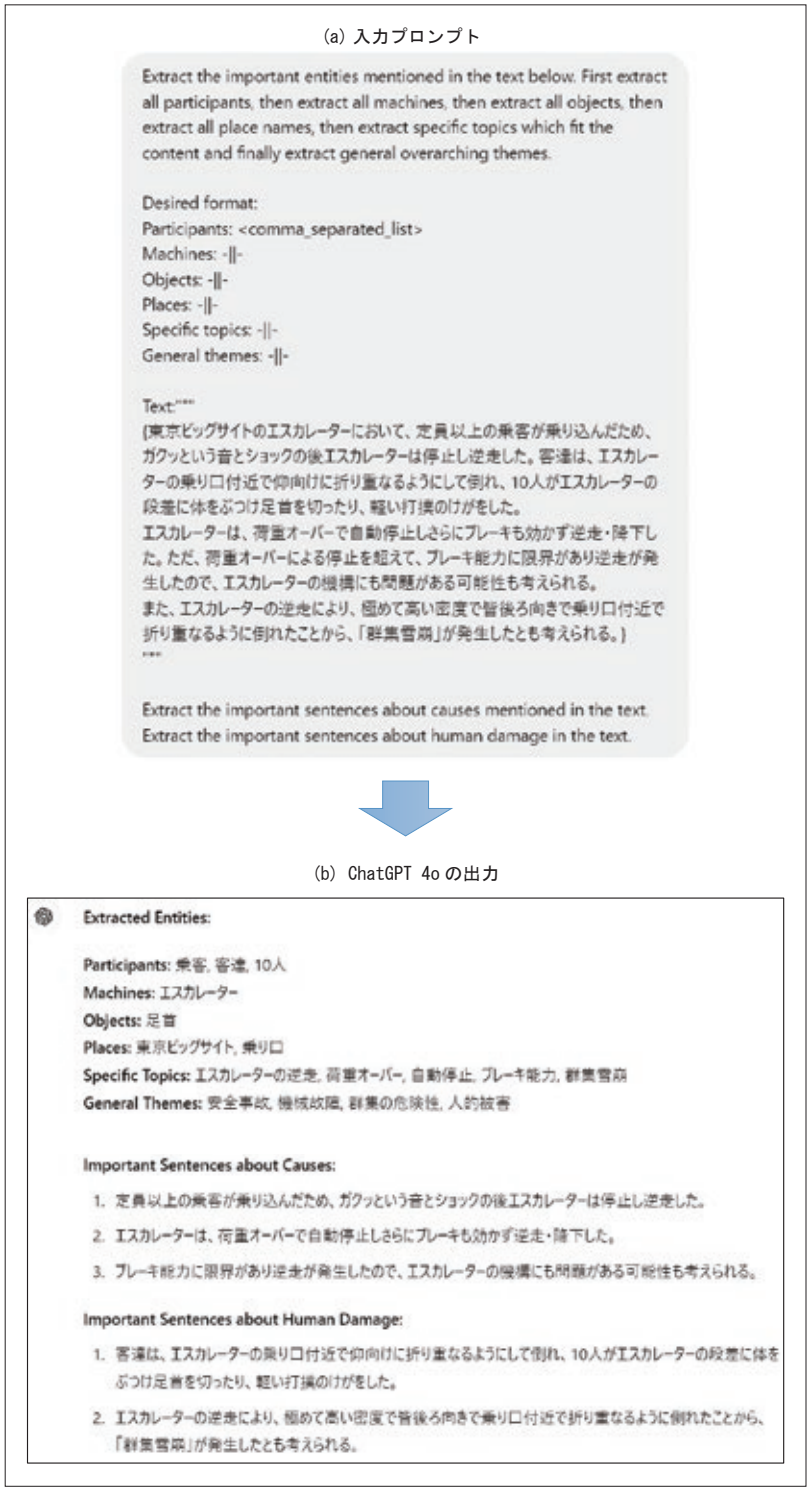
6. ハルシネーションと検索拡張生成（RAG）

表2の最後に紹介した Perplexity AI（パープレキシティ・エーアイ）はインターネット検索に強みを持つシステムで、情報源を明示しながら回答してくれる。最近の対話型テキスト生成 AI の多くは、このように Web 検索と連携する仕組みを備えている。ChatGPT も有料版では同様の機能がある。これは、テキスト生成 AI の課題である、学習したテキストには無い架空の「お話」を生成してしまうという現象、いわゆるハルシネーション（Hallucination：幻覚）を減らすための方策の一つである。

表 2　対話型の仕組みを備えたテキスト生成 AI

提供者	サービス名	URL	備考
Google	Gemini	https://gemini.google.com/	言語モデルは PaLM2, Gemini
Microsoft	Copilot	https://copilot.microsoft.com/	言語モデルは GPT-4
Perplexity AI	Perplexity	https://www.perplexity.ai/	言語モデルは GPT-3.5, GPT-4，他

図 2　ChatGPT を用いたテキストの構造化事例



ハルシネーションがなぜ起こるかというと、テキスト生成 AI の本質が、入力テキストの続きを生成してくれる「小説家」であるからだ。文脈に合った語句を確率的に選んで文章を紡いでいくので、複数の情報源から得られたものをつないだ結果、架空のお話になってしまうことがある。ChatGPT が虚偽の情報を拡散しているとして、名誉棄損で OpenAI 社が訴えられるという事件もあった[6]。

では、これをどのように防げばよいのであろうか。一つの解決策は、テキスト生成 AI に自由に文章を書かせる「小説家」としてふるまわせるのではなく、明確な作業内容や処理対象を与えて「秘書」として機能させるこ

とである。ここで Web 検索との連携が登場する。利用者の質問に基づき Web 検索を行ったうえで、テキスト生成 AI には、得られた検索結果のテキストを処理対象として要約せよ、と指示する。処理内容が要約であれば、情報源の引用ができ、検索結果のテキストの内容から大きく逸脱することもない。このように、情報検索により得られたテキストに基づいて生成を行う仕組みは、検索拡張生成（Retrieval Augmented Generation）と呼ばれる。

7. テキスト作成者は AI か人間か？

あるテキストがテキスト生成 AI によって作成されたものかどうかを見分けたいと考えたことはないだろうか。大規模言語モデルの進化により、機械が生成する文章が非常に流暢になり、人間が執筆するものと区別がつきにくくなっている。この課題に対処するために、与えられたテキストがテキスト生成 AI によって生成されたものか否かを識別するサービスが登場しているので、**表3**にいくつか紹介する。

各システムに関する説明では、人間が作成した文章と、テキスト生成 AI が生成した文章とを学習して、判別するとされているが、その詳細については明らかにされていない。文章のトーンやスタイルの不一致、複雑さ、典型的な人間の文章のパターンからの逸脱などを分析しているとの説明もある。

GPTZero については、統計的手法も援用されていると公式記事で説明されており、perplexity と burstiness に注目しているとのことである。perplexity という言葉が再び登場した。先ほどは会社名やサービス名の一部であったが、今度は、本来の意味で使われている。perplexity は、言語モデルの「悪さ」を表す尺度で、得られた言語モデルが、実際の文章における語の並びを予測できない度合いを表している[2)]。値が低いほど良い言語モデルと判断され、言語モデルで予測しづらい、多様性を持つテキストでは値が高くなる。さて、テキスト生成 AI により生成されたテキストの場合、大規模言語モデルの perplexity は、人間が作ったテキストを対象とする場合に比べて、低い値になる傾向にある。大規模言語モデルに従って語句を選んで生成したテキストなので、言語モデルの予測が当たりすぎる、つまり、良すぎるわけだ。このように perplexity の値の大小をみると、文章が人間によるものかテキスト生成 AI によるものかの目星をつけることができる。次に、burstiness についてである。日本語では、そのまま「バースト性」と表される。一般的には、時間経過のなかで、ある出来事が集中して起こる傾向を指す。ここでは、対象テキスト全体にわたって、そこに現れる特定の執筆パターンや perplexity の変化の度合いを見ている。人間による文章の場合、執筆パターンや perplexity の値が変動しバースト性を持つ傾向にある。一方で、テキスト生成 AI による文章では、言語モデルに従って規則的に生成されているので、そのようなバースト性が低いとされる。

表 3　AI 生成テキストの判別を行うサービス

提供者	サービス名	URL
GPTZero	GPTZero	https://gptzero.me/
Copyleaks	AI Content Detector	https://copyleaks.com/ai-content-detector/
Somodin	AI コンテンツ検出器	https://smodin.io/ja/ai%E3%82%B3%E3%83%B3%E3%83%86%E3%83%B3%E3%83%84%E6%A4%9C%E5%87%BA%E5%99%A8

上記のような手掛かりに基づいて、テキスト生成 AI によって作成した文章を識別する方法が提案され、サービスも提供されてはいるものの、依然として難しい課題であることに変わりがない。誤判定の可能性を考慮にいれて、利用する際には慎重に判断すべきである。

8. おわりに

本稿では、テキスト生成 AI の仕組みを平易に解説しつつ、その利活用に関わるいくつかの話題を提供した。テキスト生成 AI の存在は、我々の行動に良くも悪くも影響を及ぼし続けることであろう。大規模言語モデルで何ができて、何ができないのかは、まだはっきりわかっているとは言い難い。そのため、効果的な使い方ならびに諸問題点を積極的に蓄積し、共有することが重要であろう。

参考文献

1） Alec Radford, et al.: Improving Language Understanding by Generative Pre-Training. OpenAI, 2018.
2） Christopher D. Manning, Hinrich Schütze, 加藤恒昭, 菊井玄一郎, 林良彦, 森辰則訳：統計的自然言語処理の基礎. 共立出版, 2017.
3） Ashish Vaswani, et al.: Attention Is All You Need. In Proc. of 31st Conference on Neural Information Processing Systems (NIPS 2017), 2017.
4） Colin Raffel, et al.: Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21, Issue 1, 5485–5551, 2020.
5） 畑村洋太郎, 中尾政之, 飯野, 謙次：失敗知識データベース構築の試み, 情報処理 44（7）, 733-739, 2003.
6） Siladitya Ray.: OpenAI Sued For Defamation After ChatGPT Generates Fake Complaint Accusing Man Of Embezzlement. Forbes, 2023年6月8日, 2023.

もり●たつのり

1991年横浜国立大学大学院工学研究科博士課程後期修了。工学博士。同年、同大学工学部助手着任。現在、同大学大学院環境情報研究院教授。この間、1998年2月より11月まで Stanford 大学 CSLI 客員研究員。自然言語処理、情報検索、情報抽出などの研究に従事。