日本型セルフケアへのあゆみ　第 21 回

ChatGPT の可能性と危険性：ハルシネーション問題

児玉龍彦

東京大学先端科学技術研究センターがん・代謝プロジェクトリーダー
日本セルフケア推進協議会業務執行理事

医学のあゆみ　別　刷

Vol. 286 No. 9 2023. 8. 26

日本型セルフケアへのあゆみ

児玉龍彦
東京大学先端科学技術研究センターがん・代謝プロジェクトリーダー
日本セルフケア推進協議会業務執行理事
日本在宅がん療養財団代表理事

人生において，元気でいることは誰にとっても大事なことである．自分の健康と病気に関わることは正確に知りたい．さまざまな薬や治療法があるなら，自分の希望で決めたい．そうした願いをもとに，大きな転換がはじまろうとしている．インターネットの普及により，医薬品・健康食品・病院に関する情報に誰でも容易にアクセスできるようになったが，正確性に欠けた情報も溢れかえっている．本シリーズでは，地に足をつけた“日本型セルフケア”へのあゆみを提唱していく．

第21回 ChatGPTの可能性と危険性：ハルシネーション問題

POINT

- **2022年11月，OpenAI社は，幅広い分野の質問に対して会話文で回答してくれる人工知能チャットボット“ChatGPT”を公開し，世界中に大きな衝撃を与えた．**
- **その回答の精度の高さも注目を集めたが，“ハルシネーション(幻覚)”とよばれる嘘の答えも含まれるので，医療・金融・法律などの分野では大きな問題となっている．**
- **ハルシネーション問題の対策として，チャットボットを的確に使用するためのプロンプトを開発・最適化する“プロンプトエンジニアリング”という新しい学問分野が注目を集めている．**

生成AIの勃興

2022年11月，OpenAI社は生成AIサービスのChatGPTをリリースし，世界に大きな衝撃を与えた．言語を問わず，ブラウザ上で質問を入力すると，数秒のうちに自然な言葉で回答が得られるこのサービスは瞬く間に広まった．さまざまなジャンルの問いにももっともらしい答えを提供してくれ，まるでその分野に精通した人と会話しているような不思議な気分になる．しかもその正確さは日を追うごとに改善されている．2023年3月には，さらに進化した大規模言語モデル「GPT-4」が登場し，以前みられた間違いがぐっと少なくなり，使い勝手がよくなってきている．

ChatGPTとはOpenAIが開発した大規模言語モデル(large language models：LLM)を基にする対話型AIである．膨大な文章を読み込ませたファウンデーションモデルを使うことにより，これまでのAIの印象を一変させた．従来のAIは，まず大量のサンプルデータを読み込ませてトレーニングを行ってから特定のタスクのみを実行できるものであったが，これに対して生成AIは大量のデータを自ら学習し出力するため，文章や画像など以前は扱いにくかったクリエイティブな分野にも応用が可能となっている(**図1**)．OpenAI社以外にも，Googleの「Bard」，Amazonの「Titan」，イスラエルのAI21 Labs社の「Jurassic-2」など，多くのテック企業がこの分野に参入している．

また画像生成の領域でも，OpenAI社の「DALL-E(ダリ)」や，ミュンヘン大学の研究グループが開発した「Stable Diffusion」など続々と展開されている．これらは，数十億のWebページをスクレイピング(抽出)して画像の収集を行い，それらのデータベースをもとに，入力された情報に合った画像を自動で生成する．条件を追加すれば，絵画・写真・アニメ風などタッチも含めて指定することができる．

その手軽さと無料で使えることから，こうした生成AIサービスは爆発的に普及し，専門家のみならず一般層にも浸透することとなった．

ChatGPTの仕組み

まずは代表的なChatGPTの成り立ちについてみてみよう．「GPT」とは“Generative Pre-trained

©医歯薬出版　無断複製・公衆送信は著作権侵害です

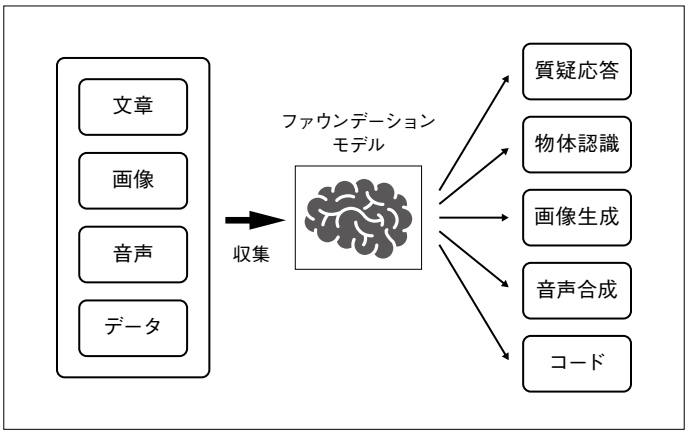

図 1　ファウンデーションモデルとは？

Transformer”の略であり，それぞれ次の意味を示している．

Generative：「生成する」という意味．既存のものではなく，自ら新しい文章や画像を生み出すといった創造性を指す．

Pre-trained：「事前学習済み」という意味で，大規模なデータセットでトレーニングされた状態を指す．このデータのなかには Web 上から自動で収集された文章・画像も含まれる．

Transformer：Google が 2017 年に発表した深層学習モデルである．連続した単語で文章が成り立つように，連続したデータの関係を追跡することによって文脈や意味を学習するニューラルネットワークを指す．

昨今の AI ブームの基盤となっているのがこの“Transformer”である．Transformer が登場する以前，ディープラーニングの分野で最も一般的なモデルは畳み込みニューラルネットワーク(CNN)や回帰型ニューラルネットワーク(RNN)であったが，ここ 5 年間で Transformer に取って代わられつつある．

Transformer は，データの関係性を学習し決定する“Attention”とよばれるメカニズムが使用されている．Attention は，文章のような前後の並びが重要なデータを扱うことにできることが特長である．従来のモデルは文頭から順番に単語をひとつずつ処理していたのに対して，Attention は文章内の単語間の関係を学習し，重要な情報を特定する．これにより，Attention を用いた Transformer は従来よりも高速な処理が可能となった．元々は自然言語処理の領域で活用されていたが，画像・オーディオ・ビデオなど，より大きなデータを扱う領域への応用も期待されている．

ハルシネーション問題への対応
――進む実証試験

ChatGPT の隆盛を受けて，医療・医学の分野でも生成 AI を活用できないか，大きな期待が持たれる．すでに画像診断の評価では AI の助けは必須になりつつあり，内視鏡の画像でも AI の助けによるがん診断率の改善が保険適用になろうとしている．ゲノム診療のようなさらに膨大なデータ処理が前提になる時代になると AI の助けは必須になっていく．

ところが，実際に ChatGPT を使って質問してみると，まったく勘違いの答えが返ってくることがしばしばである．これは，GPT がさきほど説明したように，本質的な内容を理解したわけではなく，文脈からもっともらしい答えを推測して返答しているに過ぎないからである．

GPT-2，3，4 と進歩するにつれ，正解率(その時点での理想的な答えに合致する)が上がっている(**図 2**)[1]が，それでも現状では最低 20%は誤った回答がでる．OpenAI 社のテクニカルレポートでは，正誤がはっきりしている問題で起こる間違いを“ハルシネーション(幻覚)”とよび，注意を

©医歯薬出版　無断複製・公衆送信は著作権侵害です

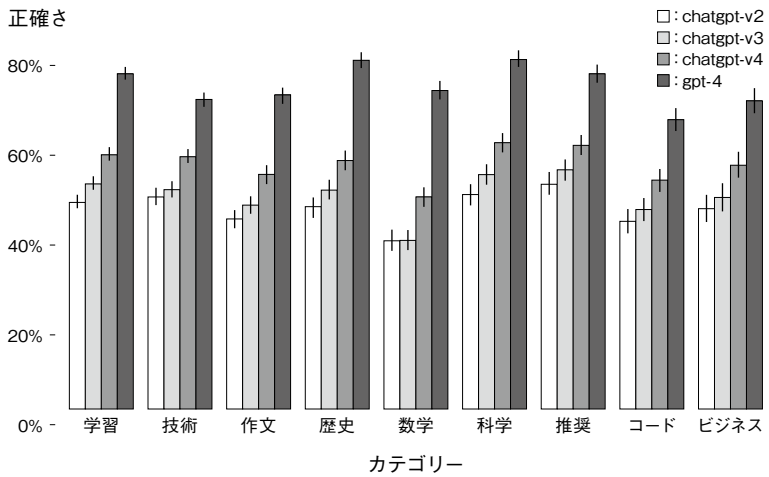

図 2　GPTのバージョンアップによる正誤率の改善[1)]

よびかけている．文章が滑らかだからこそ，誤情報が含まれていることに気づきにくい．

技術評価書で見ると，すべてのカテゴリー，特に科学や歴史，学習のカテゴリーでは，パラメータを増やして学習させると，正解率が上がっていることである．つまり，その分野に特化した大規模言語モデルを構築することで，生成 AI の正解率は飛躍的に向上する可能性がある．特に，司法試験や医師国家試験では，過去の問題にはどう応えると正解とされるかがよく研究されているので，機械的に勉強させやすい．

すでに GPT-4 は，米国統一司法試験(Uniform Bar Exam：UBE)で受験者の 90%を上回るスコア(トップ 10%)を叩き出した．日本でも，東北大学などの国際研究グループが GPT-4 に医師国家試験の問題を解かせたところ，合格ラインを超えたことが話題になった[2)]．ただし得点は基準値よりもかなり低く，また「安楽死を勧める」などの禁忌肢を選択するケースもあったと報告されている．医療の分野に絞って大規模言語モデルをより充実させていけば，理論上は，誤答率は下がっていく．現在は禁忌肢を選ぶ可能性もあり実用にはほど遠いが，生成 AI による医療・健康相談の有用性は今後高まってくるであろう．なお，米国医師会の雑誌では，患者からの質問に医師が答えるよりも AI に自動的に答えさせたほうが患者から高い評価を得ていることが，カリフォルニア大学サンディエゴ校の研究者から報告されている[3)]．

実地医療においても，国家試験に受かることは医師としての技量を何ら保証するものでなく，実際の臨床研修をはじめる資格を得たにすぎない．そこから，医師は2つの方向で経験を積んでいく．

第 1 には，試験には出ない，例外的な経験を積んでいくことである．国家試験の多くの問題は「標準治療」とよばれる，一般的な患者に行われることが推奨される治療を選択する問題である．しかし実際の診療を重ねるにつれ，「一見○○に見えるが，実は□□だった」という落とし穴のような症例の経験を積んでいく．生成 AI もデータを増やしていけば，こうした標準治療からの例外事象の可能性を見つけてハルシネーションを減らす学習を積んでいける．

第 2 は，患者の性格や人生観を理解し，その希望の実現をお手伝いする，という人間理解の方向である．たとえばがん療養においては，単なる延命よりも，苦痛や不快感を緩和し，精神的な平穏や残された生活の充実を希望する患者も増えてきている．医師と患者の間で，こうしたゴールラインが一致していないケースも多々ある．患者の本心は，医師以上に患者に接する時間の長い介護・看護の従事者のほうが先に気づくことが多い．医師は患者やこうした医療従事者の方々との交流を通じて，病気ではなく人をみる医療のあり方を学んでいく．こうした健康観は十人十色であり，

©医歯薬出版　無断複製・公衆送信は著作権侵害です

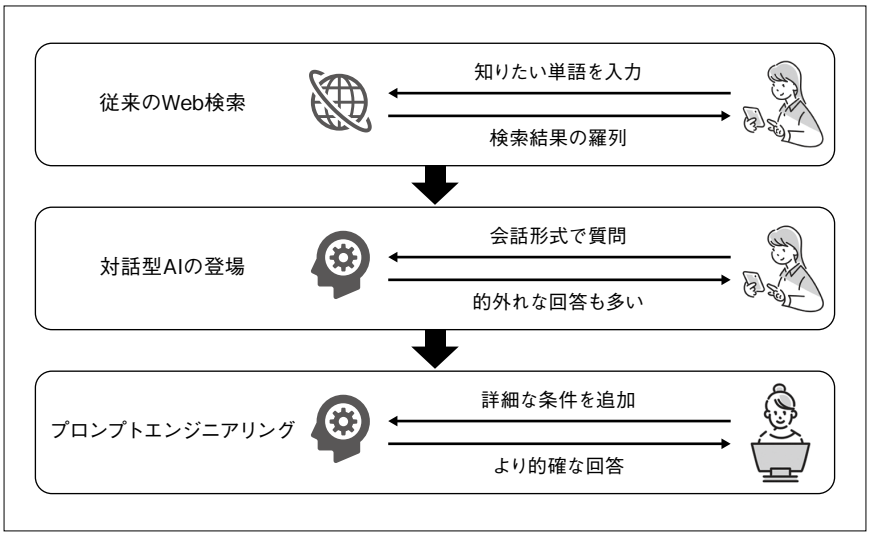

図 3　対話型AIの発展とプロンプトエンジニアリング

データを蓄積しても最適解が出せるようなものではない．診断・治療以上に AI の応用が難しいように思われる．

プロンプトエンジニアリング

生成 AI を活用するテクニックとして，プロンプトエンジニアリングという技術が注目されている(**図 3**)．

対話型 AI に質問する際に入力する文字列のことを「プロンプト」と呼ぶ．プロンプトエンジニアリングは，AI に対して適切な質問や指示を与えることで，より望ましい結果を引き出す技術である．質問の目的や，回答者(AI)の職業，質問者の知識の度合いなど，詳細な設定を追加していき，回答を目的に適ったものへとブラッシュアップする．

画像などの生成 AI を使用する際にも，使用目的は何か，写真か絵画かイラストか，状況や雰囲気を細かく伝えるプロンプトエンジニアリングが，精度の高いアウトプットを得るためには有用である．

医療現場における音声認識技術の活用

医療現場での生成 AI の応用は，入力した文章への回答だけでなく，画像や音声の取り込みも含まれる．

米国の Microsoft 社は 2021〜22 年に，2 兆円をかけて音声認識の技術を持つ Nuance 社を買収した．同社の音声認識技術は iPhone の Siri などでも使われている．Nuance 社は，スキャナーやシンセサイザーなどを開発した Kurzweil Computer Products 社をルーツとする．同社創業者のレイ・カーツワイルは，機械が人間の知力を超える技術的特異点(シンギュラリティ)が 2045 年に訪れるという説を提唱したことで知られる，人工知能の世界的権威である[4]．

1980 年代から，カーツワイルは言葉を聞いて理解するメカニズムに興味を持ち，単語・文法だけでなく，状況や，何に注目(attention)するかで意味合いが変わってくることに着目し，音声認識・解析の基礎となる考えを提示した．

会話の行われている周りの状況(コンテクスト)によって語彙や意味が変わることは，いまだに難題である．たとえば医療現場で使われる「背部痛(はいぶつう)」という単語は，非医療従事者には馴染みがなく，一般的な音声入力システムでは「廃物」などと認識されてしまう．こうしたことから Nuance 社は，特定の環境下での音声認識の精度向上に取り組んできた．同社が大きなシェアを占めるのは，放射線科の画像読影の記録である．

©医歯薬出版　無断複製・公衆送信は著作権侵害です

放射線科の医師にとって，X線やCT，PETの画像を読み解きレポートにまとめるのは時間のかかる作業である．そこで，その音声入力に特化したシステムの開発が進められてきた．

Nuance社の発表によれば，米国の放射線画像の読影に関わる医療機関の80％がNuance社の報告書作成システムを使っている．同社ではこの成功経験から，音声の認識についての技術をさらに発展させ，日本語を含めて世界の50以上の言語に応用している．さらにNuance社はコールセンター支援事業においても世界で最も多くのシェアを有するほど強く，全世界で年間80億以上の問い合わせ電話を自動応答で処理している．

こうした技術を結集し，Nuance社はMicrosoftやOpenAIと連携して，コロナ禍のなかでのオンライン診療でも存在感を強めてきた．カルテの作成は医療従事者の大きな負担になっており，診察の会話からカルテ用の臨床メモの作成をするのに4時間もかかることもあった．それが，患者と医師の会話を録音し，医療用語に特化した音声入力システムを用いて，診察後，数秒でクラウドに保存する，という展望を示している．こうしてまとめられた臨床メモは，医師の承認を受ければすぐカルテに統合される[5]．

医療者と患者・家族間のギャップ

医師などの専門家と，患者・家族などの非専門家が医療に関して話し合うとき，同じ言葉でもそれに対する受け取り方が違うことが，しばしば問題となる．筆者は，家でのがん療養に関わる情報を提供する「在宅がんウィット」[6]というWebサイトの運営に携わっている．そのサイトを通じてがんの在宅療養をめぐる質疑をしていると，医師と患者・家族間の，言葉の受け取り方のギャップに驚かされる．

たとえば，「進行がん」という用語を説明するときに，医師は「予後が悪い」「副作用が多い治療を選ぶ必要がある」など治療方針に要点をおいて説明する．しかし患者や家族は，死の恐怖や「子育てや介護はできるか」，「お金はどのくらいかかるのか」，「毛が抜けたりやつれたりして外見は変わるか」……など，日常の生活に関わることで頭がいっぱいになっていることが考えられる．このように，その人のおかれた状況や心理的背景を考慮しないことには，質問と期待される答えにズレが生じる．

AIにとって，家での療養生活や日常の健康問題に対する不安に応えることは，画像や検査などのデータをもとにした画一的な答えを提示すること以上に難題となるかもしれない．

おわりに

ChatGPTは質問者からの入力内容もデータベースとして蓄積されていく．医療にかかわる情報はプライバシー性が高く，プロンプトに入力することに抵抗がある人も多いだろう．ハルシネーションなどの問題も解決されていない現在の状況で，患者が医療情報を得るために使うにはリスクが大きい．だが，臨床メモの作成などの事務的作業にAIを活用すれば，医療従事者の時間的・精神的負担は減らせる．医療従事者は目の前の患者に向き合うことにより注力することができ，結果的に患者の満足度の向上にもつながるだろう．

AIの利便性と危険性の双方を理解し，その限界も把握したうえで，日常の診療に活用されていくことを期待する．

文献/URL

1) GPT-4 Technical Report.（https://cdn.openai.com/papers/GPT-4.pdf）
2) 東洋経済ONLINE．ChatGPTと違う？「GPT-4」使ってみたリアルな感想．2023年3月17日．（https://toyokeizai.net/articles/-/660082?page=2）
3) Ayers JW et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023; e231838.
4) レイ・カーツワイル．シンギュラリティは近い［エッセンス版］人類が生命を超越するとき．NHK出版；2016.
5) Microsoft. Breaking new ground in healthcare with the next evolution of AI. Mar 20, 2023.（https://blogs.microsoft.com/blog/2023/03/20/breaking-new-ground-in-healthcare-with-the-next-evolution-of-ai/）
6) 在宅がんウィット．（https://ganwit.jhocc.jp/）

©医歯薬出版　無断複製・公衆送信は著作権侵害です