6 人とAIの役割の境界
～AIはどこまで担うのか～

山田 誠二
YAMADA Seiji
神奈川大学 情報学部 システム数理学科／教授

AIは私たちの生活に急速に浸透している一方で、明らかに誤った回答を行うなどの問題点も顕在化している。そこで、生成AIの仕組みと限界を踏まえ、人間との違いや役割分担などの境界に着目し、どうすれば有効に活用できるのかについて改めて考える。

はじめに

現在、ChatGPTをはじめとした生成AIが日常生活に浸透し、日頃の調べもの、文章作成、画像・イラスト作成などの広範囲にわたり生成AIを活用する人が後を絶たない（**図1**）。しかしながら、この現象は、**インターネット検索**がエンドユーザの日常タスク処理（特に、調べもの）に与えたインパクトと比較してどうなのか、という学術的な問いには議論の余地が残る。また、ほとんどの生成AIは、インターネットを介してサーバで処理されるため、生成AI自身がインターネットを使っている。

本稿では、このような背景から、生成AIの仕組みと人間との違い、タスク構造を説明して、本当にAIが役に立つユーザ像について考察する。

図1　日常的なAI活用構造

まず、生成AIとは何かに対する説明を**図2**に示す。この説明自体、ChatGPTに質問した答えだが、なかなか的を射た説明であることがおわかりになるだろう。

さて、生成AIは一体どうやって、この極めて自然な言語による質疑応答・対話を実現しているのだろうか。実際には相当な作り込みがされていると考えられるが、関係者以外の我々にわかっているのは、**図3**のようなことである。まずwebや書籍などから収集した膨大な文章を入力として、**Transformer**という機械学習アルゴリズムが、単語の系列から次の出現単語を確率的に予測するための**N-gram**と呼ばれる知識に類似した知識**LLM（大規模言語モデル）**を学習し、そのLLMを用いて入力文章の次に現れる単語を予測する。つまり、N-gramは、文章などのN個の単語系列を入力として、次の（N+1）番目の単語を予測する。この予測には、どの単語がどの確率で出現するのかという出現確率が付随している。例えば、[彼、は、テレビ、を]という単語系列が入力された場合に、次の出現単語は、“観ていた”<0.7>、“消した”<0.2>、“買った”<0.1>という感じである（< >内が出現確率）。専門的には、Transformerは、**自己注意**（self-attention）と呼ばれる、次単語の出現確率に影響を与える可能性のある文脈を網羅的に探索・学習するメカニズムや単語が**埋め込みベクトル表現**されているなどの単語予測の精度を上げるための様々な工夫が凝らされているが、基本的には**図3**のような仕組みと考えられる。

生成AI（Generative AI / 生成型AI）とは、文章・画像・音声・動画・コードなどの“新しいコンテンツ”を自動で作り出す人工知能のことです。従来のAIは「分類」「予測」「認識」が中心でしたが、生成AIはゼロから新しいものを作る能力を持っています。

図2　生成AIの説明

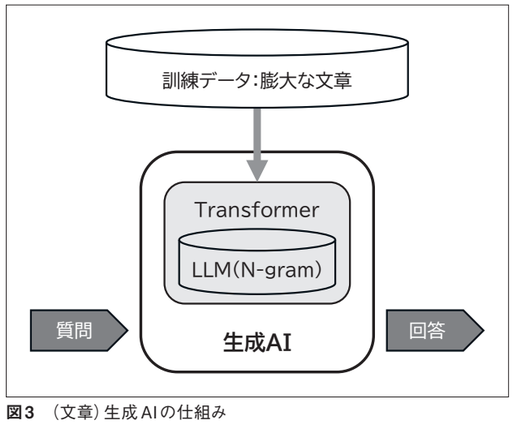

図3　（文章）生成AIの仕組み

さて、**図3**を見て、我々人間も人と対話するときや質問に答えるときに生成AIと同じような次単語予測を使って発話しているのかと自問していただきたい。確かに、挨拶や決まり切った言葉のやり取りなどのルーチン化された日常的な対話なら、LLMと同様に反射的に反応・発話していると感じるが、自分独自の意見を初めて人に伝えるような状況、**言語化されていない何かを言語化して人に伝える状況**など、どうもLLM的な処理では対応できないと感じる。そして、もっと概念レベルで（ネット上には記述されていない）多様な文脈を考慮したロジックを構築してから、そのロジックを言語化するという対話のケースは枚挙にいとまがないだろう。また、LLMには、何のためにその発話をするのかという**意図や目的が明確に記述されていないし、用いられていない**。

さらに、生成AIはあくまで次単語を予測して、文書を生成しているだけなので、「リンゴ」の説明文を生成することはできるが、「リンゴ」という**概念を「理解」しているとは言いがたい**。人間の理解の実体を直接観測することは不可能なので、「説明文の生成可能」=「理解している」という、ある種のプラグマティックで工学的な立場もあり得るが、正直、人間の直感にはそぐわない。

もう一つの問題としては、生成AIは、確率的な文章生成とはいえ、基本的には、ネット上を中心とした多数派の意見に強く引っ張られる傾向がある。つまり、プロンプトの書き方にもよるが、生成AIからは、**メジャーで当たり前の回答**しか返ってこない。

このような生成AIの限界が、優れた生成AIでも20%を超えると言われる**ハルシネーション**（明らかな間違いを平気で回答する現象）や**AIスロップ**（生成AIのゴミのような回答）の発生要因となっている。

生成AIは、このような問題を抱えつつも、極めて自然な言語表現と模範的な答えの観点で、従来の対話システム・QAシステムをはるかに凌駕（りょうが）している。さらに、自然言語による対話・QAという誰もがすぐ試せるキラーアプリを実装していることも相まって、世界中で爆発的な数のユーザを獲得している。

タスク構造：リライアンスとコンプライアンス

人間とAIが協働する系で、人間とAIの役割と記述したものが**タスク構造**であり、典型的な2つのタスク構造が、**図4**に示す**リライアンス構造**（**図4（a）**）**とコンプライアンス構造**（**図4（b）**）である。

リライアンス構造は、あるタスクを人間が自分でやるか、AIに任せるかを、人間が決定し、それを実行する。AIに任せた場合、AIの答えをまったくチェックしない。これは、チェックすると効率が上がる保証がなくなるからである。リライアンス構造は、ある意味AIに全幅の信頼を置いて、任せっきりにしている。**自動車の自動運転**は、典型的なリライアンス構造と考えられる。自動運転のように、AIの出した答え（ステアリング）を人間がチェックする暇がないことも大きな理由となる。また、このタスク構造は、AIがほとんど間違わないという前提が必要となる。

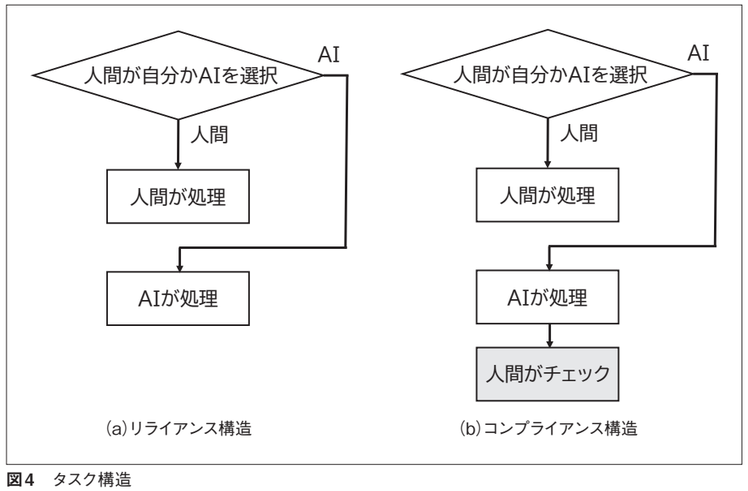

図4　タスク構造

一方、**コンプライアンス構造**では、AIに任せた場合でも、**AIの答えを人間がチェックする**（**図4**(b)のグレー箇所）。例として、よくある生成AIの利用場面では、まず生成AIにプロンプトを入力して、解（文章や画像）を生成させ、その結果をチェックして、満足できなければ、修正したプロンプトを考えてまた入力するということを満足いくまで繰り返す。このように、我々が日常的に生成AIを利用する場合の多くは、このコンプライアンス構造（の繰り返し）となっている。しかし、このコンプライアンス構造では、リライアンス構造に加えて、**チェックするコスト、プロンプトを修正するコスト**などが発生し、これらのコストの大きさによっては、かえってコスト高になってしまい、「最初から自分でやっておけばよかった！」となる場合がある。

リライアンス方程式

図4の「人間が自分かAIを選択」（ひし形部分）という意思決定で、人間が考えるべき方程式は、リライアンス方程式と呼ばれ、下式で表現される。

$T_H \leqq T_A$　T_H：人間の成功確率、T_A：AIの成功確率

この式は、「AIを使う費用対効果の計算式」とも言える。この方程式（正しくは、不等式）は、人間がタスク処理を行った場合に成功する確率である**人間のタスク成功確率**T_Hと**AIのタスク成功確率**T_Aの大小比較をしている。もちろん、タスク成功確率が高い方を採った方が系全体のパフォーマンスは大きくなる。なぜなら、タスク処理を繰り返す場合、人間とAIのうち、タスク処理が成功する可能性が高い方が常に選ばれるので、その成功回数が最大化されるからである。

図5にタスク処理を繰り返し、途中でAIのタスク成功確率が人間のそれより低下してしまった場合のパフォーマンスのグラフを示す。このグラフでは、青線が人間のタスク成功確率T_H、赤線がAIのタスク成功確率T_A、緑線がリライアンス方程式で大きい方を選択した場合のタスク成功確率を表している。この3つの曲線の下の部分の面積が（累積）パフォーマンスを表すため、リライアンス方程式による戦略（緑線）が下の面積最大となることがわかる。なお、この図にあるような、AIのタスク成功確率が人間より低下するという現象は、環境変化などにより、普通に起こり得る。

AIが本当に役に立つユーザとは

以上の議論を踏まえて、どのようなユーザが（生成）AIを利用した場合に、本当にAIが役に立つのか（業務効率が向上するのか）を考察する。結論から言うと、「連続するタスク処理において、常に自分とAIのどちらがやった方がいいかを考え、いい方に任せることができるユーザ」ということになる。換言すると、**図5**の緑線の戦略を採ることができるユーザとなる。

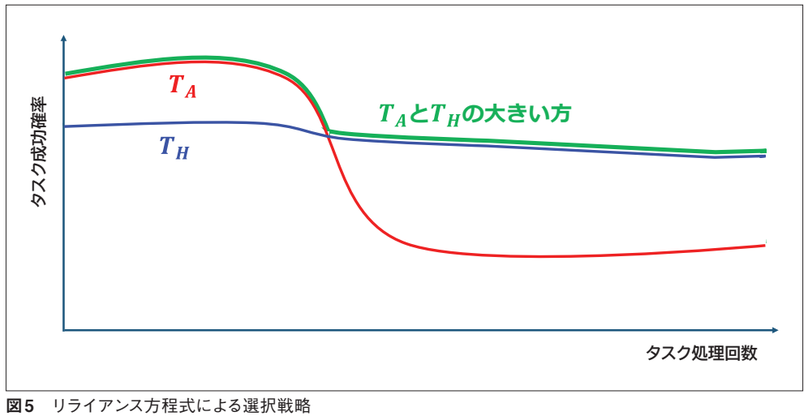

図5　リライアンス方程式による選択戦略

しかし、一般には、この人間（ユーザ自身）のタスク成功確率とAIのタスク成功確率を知ることは簡単ではない。言い換えると、「経験から自分の能力を知っており、かつ経験あるいはメディア情報によりAIの能力も知っているユーザ」となる。

この条件は重要であり、日常的に意識することで、ユーザは本当の意味でのAI活用が可能となるであろう。

人とAIの役割の境界

ここでは、人とAIの役割の境界について考察する。まず確認しておきたいのは、リライアンス方程式及び信頼較正自体は、人とAIの役割の境界を規定するものではなく、ユーザが人とAIの役割の境界を正確に設定し意識することで、AIを正しく・有効に活用することができ、その結果として人間-AI系全体のパフォーマンスを最大化できるということを示している点である。これは、一見当たり前のことに聞こえるかもしれないが、この当たり前のことを簡単な数式を使って定式化することで、そのことを対象化し、根拠を与えることができると考える。

以上の観点を確認した上で、ここで敢えて人とAIの役割の境界について、第3次AIブームから現在の生成AIの隆盛までのスコープを踏まえて考えてみる。まず、前述のように生成AIは、基本的に学習されたN-gramモデルにすぎないという認識から、哲学者デカルトの有名な命題「我思う、故に我あり」における「思う」を生成AIが実現しているとは思えない。若干繰り返しになるが、生成AIがやっているような「これまでの発話した言葉から次に発話すべき単語を予測する」ということを我々人間が頻繁にやっているかというと、そうではないと感じる方が多いのではないだろうか。では、何故そこに違和を感じるかというと、このLLMがやっている、次の単語の予測には、発話者の「意図」が欠落していることが大きな原因だと考えられる。人工知能研究者の立場からは、この「意図」は（非連続な）単語列で表現される「文脈」である程度は記述されていると考えるかもしれないが、膨大な高次元配列であるLLMから、その意図を抽出するのは、ディープラーニングの学習モデルからある概念を取り出すのと同じかそれ以上に難しい。つまり、この発話者の意図の生成（換言すると、目的・目標の生成）が、人とAIの役割の境界を決定する一つの重要な要因になっているのではないだろうか。なお、通常、人間のユーザは、生成AIに対して、プロンプトの形で自身の意図を伝えており、AIが自身で意図を生成して動くわけではない。そして、この意図（目的）は、「個体保存」という究極の意図（目的）を持った生命体にしか生成できないものではないだろうか。

まとめ

本稿では、生成AIの仕組みを概観し、人間との違いを議論した。続いて、人間-AI協働系における人間とAIの関係を表すタスク構造を紹介し、人間とAIのタスク成功確率とそれらを用いたリライアンス方程式について説明した。さらに、本当にAIが役に立つユーザの条件と人とAIの役割の境界について、筆者なりの考察を展開した。本稿が、一般のユーザが生成AIを日常的に活用する際の指針の一つになれば幸いである。