生成AIは英語論文執筆の救世主か悪魔か？

坂　野　正　明
〈Wise Babel Ltd（@WiseBabel）, Scotland, https://wisebabel.com〉
e-mail: sakano@wisebabel.com

生成AI型チャットボット，すなわち自然言語処理型汎用AI（以下，AI）は近年長足の進化を遂げ，今なお急速に進化中である．本稿では，科学編集及び英文校閲サービスを長年にわたり専業としてきた筆者が，現在のAIの一般的特徴をまとめ，特に非英語ネイティブスピーカーの科学者が論文執筆するにあたってAIをどう使うべきか，あるいは何を避けるべきか，AIの長所とリスクとを天秤にかけながら提案する．AIは正確さや真実を犠牲にしても「統計的により多くのユーザーをより喜ばせる」ことに最適化されているため，科学の分野では不都合な側面があることは確かながら，注意深く使えば十分有用である．人間ユーザーがAI出力を吟味できるだけの能力を持つことが，現代，そして近い将来においてかつてないほど決定的な要素になると主張する．

1.　計算機による自然言語処理とAI

生成人工知能（AI）型チャットボット，すなわち自然言語処理型汎用AIの近年の急速な進化は目を見張るものがある．自然言語処理は，計算機ソフトウエア開発の現場で長らく鬼門だったと認識している．自然言語はどの言語も例外が数多い上，一つの単語が複数の意味を持つことも普通だ．たとえば英文「Can can can a can.」（キャンさんは缶を缶詰にできる）を正しく解釈するための自然言語処理など，プログラマーにとって悪夢だ．さらに悪いことに，文脈次第で同じ文面の意図が異なることさえある．たとえば英国英語の「Would you like another cup of tea?」は，親切に茶のおかわりを提供する意図の時もあれば，「そろそろ帰ってくれませんかね？」の意味になることもある．ネイティブスピーカーの人間であれば状況に応じて妥当な解釈を高確率でできるだろうが，ほぼ無数のパターンが考えられる状況ないし**文脈**を伝統的プログラムで逐一判断させるのは極めて困難だ．しかも，どういうパターンがあるかは世界最大の辞書でもごく一部の例しか記載されていない．つまり言語化されていない．だから，言語処理のための「ルール」を参照する先すらなく，自然言語処理を目指すプログラマーが自力で見つけ出す必要があることになる．

また，人間は言語表現を誤る．たとえば「遍在」と書くつもりが「偏在」と誤記して正逆の意味になってしまうなどのミスはありがちだ．人間の読者ならば相当程度，**文脈**から判断できようが，伝統的計算機処理は**文脈**判断に向いていない．

特に日本語の場合，欧米言語のように単語がスペースで区切られないため，表記上の境界が曖昧であることも，伝統的な計算機処理とは一層相性が悪い．

かように自然言語処理には原理的困難がつきまとうことが長らく常識だったと理解している．2022年11月，生成AI型チャットボット（以下，チャットボット）であるChatGPTが登場し，世界中のユーザーに無料ないしそれに近い費用で利

用可能になったことは，革命的だった．人工知能（AI）と呼ぶにふさわしい，あるいはそれに近いものが広く一般人に利用可能になったイベントとして，記憶に新しい．

3年あまり経った今では，Google社Geminiはじめ，同種の優れたチャットボットが何種も公開されている．これらチャットボットの多くは複数言語に対応していて，自然言語でファジーな質問（プロンプト入力）を行っても，そう悪くない回答が返ってくることが多い．半ば当然に，校閲や翻訳にも相当の力を発揮する，とされる．それを科学論文にも応用できないか，と考えるのはごく自然なことであろう．

本稿では，まずチャットボットの原理をおさらいすることで，どういう用途が有用かあるいは不得手かの見当をつける．ついで，研究者が直面する現実的ユースケースについて，現在のチャットボット一般の特徴，特に欠点をまとめる．その際，原理的なポイントに焦点を当てて，チャットボットの種類や近未来の進化に依らないなるべく普遍的な特徴に絞っている．

そのうえで，現在あるいは近い将来，研究者のユースケースにおいて，チャットボットが有用なケース，要注意なケースとリスク，そして人間がやるべきことについて，提案する．

現代は，研究者あるいは広い意味では知的労働者層全体が，人類が未だ経験したことのない急速な変化の中にいることは間違いない．AIシンギュラリティが訪れたならば，自分たちの存在意義さえ問われかねない．専門文書の科学的編集および英文校閲を生業とする筆者は，現在その荒波の最前線に立つ立場であり，大いに危機感を募らせている．本稿の末尾には，そんな筆者から，特に将来ある若年層へ向けた長期的視点からの提言も行っている．

2. AIチャットボットと機械学習

2.1 機械学習

チャットボットの本質は「学習」（機械学習と呼ばれる）にある．人間は，幼少の頃から大量の情報を経験し記憶する中で一定の規則を（多くは無意識のうちに）見出すことで学習する．その有機的な積み重ねにより，極めて複雑な自然言語処理をもマスターするものだ．

伝統的プログラムではすべての場合分けとそれに応じた制御分岐を人間が陽にコーディングしていた． AIはそれとは対照的に，AI自身が多様な状況をいわば人間のように学習し，それに応じて場合分けと制御分岐を自発的に生成する．AIの機械学習は人間のように「規則を見出す」わけではなく，一旦学習したことは忘却せず，判断も出力も確率的，などの大きな違いはある．しかし，多様な状況に応じて多様な出力を出す，という意味では人間の判断に似ていて，多様な判断が要求される自然言語処理とも相性がよい．

AIの判断が機械学習によって形作られる以上，AIの判断（そして出力）の質は，学習及び解釈の精度に加えて，機械学習のためのトレーニングデータの量と質（精度）に強く影響される．これは人間の学習の類推で理解できる．たとえば日本語を母語とする人が英語を学ぶ時，読む英文の量が絶対的に不足していたら，あるいは大量に読んでも悪文ばかり読んでいたら，上達は限られる．

機械翻訳の場合，トレーニング・データとして（少なくともかつて）世界でもっとも著名な標準は欧州連合（EU）議会の議事録だった．EU議会議事録は，専門の翻訳家により，（2026年現在は）EUの公式言語である24ヵ国語に翻訳される．その翻訳の正確性（質）もその量も，世界最高水準だ．また，現実問題として，有料の市販小説などと異なり，データ（議事録）が無料で手に入るのも大きい．そのおかげで，EU加盟諸国の言語間の機械翻訳の精度は，それ以外の言語よりも高い

傾向があったと聞く．逆に言えば，日本語や中国語も含めたそれ以外の言語は機械翻訳の精度の発展という意味で，不利だったことになる．

言語間の距離も当然問題になる．英語を中心に考えた場合，大半の欧州言語は同じ印欧語族に属していて言語間の距離がそもそも近い．そして，アルファベットという共通文字を使う（一部の国や地方ではギリシャ文字やキリル文字など）．日本語は対照的に，英語との言語間の距離が極めて遠く，そのため，翻訳の自由度がはるかに高くなる．たとえば，英語「I love you.」の仏語訳はほぼ「Je t'aime.」で決まりのところ，日本語では，話者と聞き手の関係性や状況に応じて何十通りもの訳が考えられる．機械翻訳エンジンはそれらのパターンを文脈に応じた確率として学習する必要がある．すなわち，英日機械翻訳は，英仏機械翻訳と同程度の翻訳精度を達成するためには，ずっと大量のトレーニング・データが必要になろう．

同様のことは，翻訳に限らず，機械学習を基礎とするほぼすべてのAI出力に当てはまることだろう．たとえば，数多くの人が書くような内容，たとえばクリスマスカードの挨拶の文面，は膨大なトレーニング・データがあり，結果，チャットボットを含めて機械学習ベースの翻訳や校閲においては，かなりの高精度が期待できる．一方，世間に文献・文例が少ないマイナーな分野であればあるほど，精度が落ちると予想できる．トレーニング・データの質も量も落ちるからだ．

研究者が英語を書く場面には，論文やプロポーザルの執筆以外に，論文査読者との折衝，外国組織や外国人研究者との事務連絡，公募書類応募などがあろう．筆者の観測では，英語表現という意味では，日本語を母語とする研究者は，実は英語論文執筆よりも論文査読者との英語での折衝の方が，苦手とする傾向があったものだった．端的には，筆者が校閲依頼を受けた時，論文本文よりも査読者への返信の方が，朱を入れる箇所が多かった．返信原文には英語的に失礼なものもよく見かけたものだ．日本在住研究者は一般に，英語論文はよく読んでいて論文英語には慣れ親しんでいるのに対し，誰かと英語で準フォーマルなやり取りをする機会は限られるだろうことを考えれば，後者の方が壁が高いのは当然の傾向だと納得できる．

その点，AIチャットボットの登場は，ありがたいことになりそうだ．謝礼や謝罪や問い合わせといったビジネスレターは世界に無数に近い実例があるから，AIによる機械学習の機会もほぼ無数にある．結果，ビジネスレターは，AIチャットボットが真価を発揮するフィールドだと観察する．論文査読者との折衝の手紙の場合，肝になる内容こそ専門的で特殊ながら，人間相手の交渉としての英語表現に関しては，普通のビジネスレターに準じる．つまりチャットボットは有用だろう．

対照的に，科学論文は，使われるケースが狭い科学者の世界に限られる以上，サンプルは桁違いに小さい．加えて，多くの学術雑誌が有料であるから，AI機械学習の機会はさらに限られると予想できる．にも拘らず，専門分野ごとに特殊な専門用語や時には慣用表現が存在するので，AIといえども敷居が高くなっている，すなわちAIの手助けの精度はずっと落ちる，と予想できる．

2.2　AIへのフィードバックと強化学習

チャットボットでは，一般に，人間の一般ユーザーがフィードバックを送ることができる．AIチャットボットはそのフィードバックを使って「強化学習」を行い，以降の回答精度を高めようとすることができる．なお，ユーザーは，仮にフィードバック専用のインターフェイス（「いいね」ボタンなど）を陽に使わなくても，チャットボットとの自然言語でのやり取りの中で，人間相手に会話するのと同様に，謝礼を伝えたり不平や批判を伝えることもよくあるだろう．そういった反応もまた，ユーザー・フィードバックとしてAIチャットボットの強化学習に使われるだろう．

このようにチャットボットに誰でも正解を教える（一種のAIトレーナー）ことができると，い

わゆるクラウドソーシングのような形でチャットボットの強化学習が進んで，それは好ましいことのように一見思える．実際，たとえば「1＋1＝3」とチャットボットが誤答した時，「正解は2だ」と教えるレベルのことならば，このフィードバック学習機構は理想的に動くことだろう．ユーザーとしても，使用チャットボットの精度が向上するのは喜ばしいことなので，フィードバックを積極的に送る動機もある．

しかし実は，一般ユーザーのフィードバックには，原理的な問題が潜む．それは，ほとんどのケースにおいて，ユーザーは完全に公平中立かつ能力ある「評価者」とはほど遠いことにある．一般に，正解かどうかを正確にフィードバックできるのは，その評価者の方が知識量で圧倒している時に限られる．けれども，そもそも自分がそこまでよく知っていることについて，わざわざチャットボットに尋ねることがどれくらいあるだろうか？　現実は，程度の差こそあれ「正答を知らないから」チャットボットに尋ねるケースが大半であろう．その際，ユーザーは正答を知らない以上，チャットボットの回答が正しいかどうかをどうやって判断できようか？　回答を受け取った後，可能ならば教科書や辞書などの独立な方法で，丁寧に研究調査を行って回答を吟味精査したうえで評価するならば相当だ．しかし現実は，「チャットボットの回答が『一見』正しく（もしくは誤って）見える時」に脊髄反射でフィードバックを送るユーザーが大半ではなかろうか？　ならば，フィードバックを送っている一般ユーザーの大半は，中立の評価者として本来不適任の可能性が高いと推測できる．

加えて，人間は，心理学で言う「確証バイアス」から自由であることが難しい，という厄介な性質がある．「確証バイアス」とは，自分がある方向性の思想や考えを持っている時，補強する説や言論を受け容れがちで，逆にそれを否定する説や言論を否定しがちである傾向のことだ．結果，相反する思想を持つ2つのグループ（たとえば政治の右派と左派）を議論させると，議論の後，一般的に，議論前に比べて両者妥協した（たとえば中道的な）思想を持つに至るのではなく，むしろ双方とも自分の思想や自説により自信を深め（＝確証し）先鋭化していく（たとえば極右，極左思考になる）傾向があることが知られている（“Risk: The Science and Politics of Fear”（Dan Gardner 2008））．

確証バイアスのため，チャットボットへのフィードバックにおいては，「チャットボットの回答が自分が期待していた考えを強化してくれた時」に高評価，逆に「自分の偏見や誤解を否定された時」は低評価，を送るという偏った傾向があると予想できる．たとえば，「阪神タイガースが強いか」とチャットボットに問いかけたなら，関西であれば「強い」という回答により多くの高評価がつきやすい傾向がありそうだ．つまり，ユーザーフィードバックによって，チャットボットが偏った方向に「強化学習」されてしまう恐れがある．

科学的姿勢・思考・方法においては，観察者のバイアスを徹底排除することが基本中の基本だ．そうすることで初めて，客観的な事実をあぶり出し，説得力ある論理的演繹を行うことが可能になる．だから，科学者がチャットボットを使う際には，「可能な限り客観的に正しい回答」を求めるものだろう．しかし，チャットボットには，むしろ「より多くのユーザーをより喜ばせる回答」を生成するように最適化される傾向がある，もしくはそのリスクがある，と意識しておきたい．少なくとも，チャットボットが一般ユーザーのフィードバックを使って強化学習している限り，その傾向はほぼ不可避だろう．この2つの方向性が結局同じ回答に至るならば，何も問題ない．しかし，人は一般に，真実よりもむしろ自分の聞きたいことを聞きたがることが少なくないという現実を考えれば，この2つの方向性はしばしば一致しない

恐れがある．

また，商業的視点も無視できない．そもそも著名チャットボットを運営しているのはいずれも私企業である以上，資本主義の原理に則り，ユーザーが求めるサービスや製品を供給する企業が最も利益をあげやすい方向に動く，ないしはそういうサービスこそが最大の人気を獲得して市場の勝者になりやすい．つまり，市場を制するチャットボットは，「客観的に正しい回答」を返すものではなく，「統計的により多くのユーザーをより喜ばせる」ものになるという淘汰圧が働くと推測できる．その観点から，一般ユーザーのフィードバックを使って強化学習することは，商業的に尤もな方針であり，それに起因するチャットボット回答のバイアスも今後とも残ると予想できる．

さらに加えれば，チャットボット開発者個人の偏見が反映される可能性もある．Grok擁するX（旧Twitter）社のイーロン・マスク氏が（LGBTの概念などで）そういう方向性の発言で繰り返し炎上したことは有名だ．

2.3　チャットボットの最大の問題点

真実を求める徒にとって，現在のチャットボットの最大の問題は，「さも本当のように堂々と嘘をつく」ことであろう．同様に，不確定のことを不確定と述べることはチャットボットは苦手とする．広い意味では，それも嘘つきの一形態だ．

第2.2節で述べた「客観的に正しい回答」よりも「統計的により多くのユーザーをより喜ばせる」方の優先度が高い，という現代のチャットボットの本質を考えれば，この嘘をつく傾向は，半ば必然だろう．というのは，「さも本当のようにつかれた上手な嘘」を人は信じる傾向があるからだ．中には注意深く嘘を見破って不満を抱いたり低評価フィードバックするユーザーもいるにせよ，嘘を信じて高評価するユーザーの方が統計的に多くなりそうだ．結局，大半のユーザーは自分の知らないことをチャットボットに尋ねている以上，積極的に検証する能力と行動力がない限り，回答の嘘を見破ることもできない．

対照的に，チャットボットが「確証はないが○○の可能性がある」というような正直な回答を仮にした場合，その歯切れの悪い回答に満足して高評価するユーザーは多くなく，むしろ「うまい嘘を信じて高評価する」ユーザーより少ない傾向があるのではないか？

以上より，チャットボットの強化学習が「より多くのユーザーをより喜ばせる回答」を目指す方向性に向いている以上，チャットボットが嘘をつく傾向は，今だけでなく近い未来もなかなか改善されない，と恐れる．加えれば，嘘が見破られたら「ユーザーをより喜ばせる」目的に失敗する以上，チャットボットが嘘をつく時は見破りにくいものを選択的に選ぶように強化学習されると推測できる―性質が悪い．

同じ背景だろう，筆者の経験でも，チャットボットに日本語の報告文を校閲させた時，原文では「～らしい」「～と聞く」などと伝聞調に書いていたところの大半が，「～だ」と断定調に校閲されたこともあった．明快で歯切れよい文の方が読者にウケやすいならば，そして原稿校閲の目的が読者ウケを至上命題としているならば，チャットボットのこの校閲は一概に悪とは言えないだろう．しかし，特に正確さを重んじる分野，たとえば科学論文やジャーナリズムや法律関係の文書において，このような校閲は深刻な問題になり得る．実際，論文においては，何が確実で何が伝聞かを切り分けることは極めて重要なポイントの一つだ．

現代のチャットボットは正確さを旨としない以上，科学論文執筆との相性は必ずしもよくない．使用にはリスクが伴うので要注意，と結論できる．論文ではトレーニングデータが限られて精度が落ちる（第2.2節）ことも不利に働く．加えて日本語も絡む（日本語論文執筆やその英訳作業）場合は，トレーニングデータがさらに限られて精度が一層落ちることになろう（第2.1節）．

3. 論文執筆でチャットボットを使えるか？

科学論文へのチャットボット応用に問題があることは確認したが（第2章），リスクを理解して用途を選べば，便利なツールになり得ることは間違いない．本章では，リスクを念頭におきつつ，代表的な有用なケースを列挙する．工夫次第でこれ以外にもいろいろ有用なユースケースは考えられることだろう．たとえば，論文に含む図表の製作においても，チャットボットが役立つことは疑わない．

3.1 チャットボットが特に有効なケース

・スペルチェック．文法チェック．フォーマットの統一化（たとえば文献引用方式の統一や英国英語綴りと米語綴りの統一化）．
—従来のスペルチェッカーと異なってAIは，辞書に存在しないスペルだけでなく，文脈的に不自然な単語を，著者が意図しただろう似たスペルの単語に修正してくれることがよくある（例：meet⇔meat）．
・時制の（章内での）統一化．
・ど忘れしたことを思い出す時．
—たとえば「Wednesday」の正確な綴りが思い出せないまでも，見ればわかる，という場合．仮に「Wensday」などと回答された場合，それは誤りだと即座に判断できるだけの前提知識があることが条件．
・類語辞典（シソーラス）代わりに使う．単語だけでなく，代替表現の可能性も．
—提案された例をそもそもよく知っているという条件付きで，原稿を改善できる可能性．
・調査の時短として使う．特に，インターネット検索エンジンや文献検索インターフェイスの上位互換版として使う．
—原則はあくまで「時短」すなわち調査の時間を短縮することが主目的である．本来，ほかの方法，たとえば複雑な詳細検索条件指定してのグーグル検索や，図書館で文献読み漁る，でも解決できるところ，同じことを（自然言語検索により）「相対的に」高速に実現できる，ということが鍵．
—不正確な回答も少なくない．複数（数多く）のプロンプトを使う，従来の検索エンジンと組み合わせる，信頼度がずっと高い情報ソース（定評ある教科書など）と照合する，など，回答を検証すること．チャットボット出力は，グーグル検索での最上位1〜3個だけに絞って回答しているのに近い，と解釈していいだろう．

3.2 チャットボットに一定の意味があるケース

これに加えて，最新のチャットボットであれば，さらに一歩進んだ助言を求めることも実用的になりつつあるだろう．

・「表現の硬軟（口調）」の統一化．
—たとえば論文英語は口語表現ではなく堅い表現で統一すべき．日本語文章ならば常体（〜だ）と敬体（〜です）の統一など．
—母語の場合は自分でAI校閲の正当さを判断できるだろう．一方，外国語の場合，回答を吟味できるだけの語学力が自分にあるか？チャットボットの回答は，単に表現の硬軟の変化でなく，実は文意が変わってしまうこともある．要注意．
・あやふやな記憶を証拠とともに確認する．
—ただし，「自分の理解を裏付ける確かな証拠を得たい」とは**確証バイアス**（第2.2節）の典型的温床であることに注意．
—チャットボット回答を批判的に吟味することはほぼ必須．一つの方法は，本来の質問と正逆のケースを支持する証拠を要求するプロンプト入力すること．たとえば，「AがBである証拠」を求めた後，「AがBではない証拠」を求め，両者を比較検討する．
—昔から，何か対立する有力仮説がある時は，その両者の主張を読み込んで理解したうえ

で，公平な目で自分で判断するべきであった．AIも人間と同様に偏っている可能性がある以上，この状況は何も変わらない．

・「最適化」したい場合．

—たとえば原稿で同じことを繰り返して述べてしまっている時に，すっきりさせるため．

—前提条件として，「最適化」の出力結果を自分が吟味できる能力があること．

—「最適化」の結果，重要な情報（単語レベルからパラグラフレベルまで）がしばしば落とされることに注意！　典型的な例として，伝聞として記述したものが断定口調に修正されたり，文献が落とされることはよくある．

・論文における論理の不備を指摘させる．科学論文としての質を向上させる．

3.3　論文の日英翻訳にチャットボットを使う是非

論文執筆経験の浅い（もしくは皆無の）学生やジュニア研究者にとって，

・母語（日本語）で論文全文を書いてそれをAIに英語に翻訳させることは推奨ないし許容されるか？

この疑問自体はAI出現以前からよく議論があったものだ．当時，機械翻訳はまだほぼ使い物にならず，日本語で書いた原稿は人力で全文英訳する手間をかけざるを得なかったにもかかわらず，日本語で初稿を書く人は少なからずいたようだ．しかし，日本語に頼らず，原稿を一から英語で書くべきと主張する人もいる．筆者も，大学での講演の後に学生からその是非について質問を受けたことがある．

賛否があることは承知している．「英語で一から書くべし」派の論拠はおおむね，(1) 日本語で書いているといつまで経っても英語が上達しない，(2) 英語的思考と日本語的思考は異なるから日本語で考えて書いた論文は英語に翻訳すると質が悪い，の2点が主だと認識している．強いて付け加えれば，自分が苦労したから後輩にも苦労してほしい，という捻れた感情が背後にあるケースも存在するかもしれないが．

しかし，現実的に，今まで一度も英語論文を書いたことなく，かつ英語に苦手意識を持つ学生にとっては，初めての英語論文執筆は手がかりさえないことは大いにあり得る．そんな時，日本語で書いた初期原稿をAIに全文英訳させることが「手っ取り早い」方法であることは疑いない．原稿は，英語的思考にはなっていないだろうが，その英語力レベルの人は，当然ながら，純粋英語的思考をできはしない．だから，AIに英訳させたものを初版原稿としてそこから磨いていくのは，筆者は悪い方法だとは思わない．実際，「学生が人生で初めて英語論文を書いた時，昔ほど壊滅的なものが少なくなった」と語っていた教授もいた．AIに英訳させていたのかもしれない．

英語教育の観点からは，日本語を経ずに直接英語で執筆する経験はいずれ積み重ねていくべきではある．さもないと，英語的感覚，そして英語で論文の論理を構築する感覚，がなかなか身につかないからだ．とは言え，そのためには専門用語や業界特有の表現も含めた一定の英語能力とまた英語論文の典型的構成についての知識は必要になる．むしろ，初回は，自分の日本語を機械翻訳させた英語出力から，英語論文の典型的構成や表現を逆に学ぶこともできるだろう——ただし，不自然な訳になることもあれば，英語としてまともに成立していない文になることもあるので，決して鵜呑みにすることなく，英文法の知識をもって検討することは必須条件になる．たとえば日本語では体言止めの文がよくあるが，英語ではそれは不自然なことが通常であるため，そんな日本語原文からの直訳は酷い英文になることがある．また，そもそも日本語原文に筋が通っていないと，機械翻訳させてもまともな文になることは当然期待できない．念には念を入れて検討されたい．

また，全文ではなく，スポット的に機械翻訳を使うことは，多くの研究者に役立つ手法かもしれない．科学論文においては，しばしば論理が複雑

に入り組み，母語の日本語でさえ表現に苦労することがある．一番の推奨は論理自体を整理して明快な筋を再構成することながら，いつもそうできるとは限らない．そんな時，無理に英語を綴ろうとしても，文法さえ破綻したりして意味不明になりがちだ．筆者の英文校閲現場でもそんなケースは何度も経験ある．そういう時は，依頼者に論理関係を日本語で説明してもらって，その部分だけ校閲者の筆者の方で英語で一から作文することが多い．その役をAIに代わってもらうことは（優れた人間の翻訳者には及ばないまでも）可能であろう．

3.3.1　機械翻訳の注意点

2026年現在，多くのSNSで自動機械翻訳が実装されている．どのプラットフォームもAIを活用しているのだろう．しかし，その翻訳を見比べれば，少なくとも英語⇔日本語翻訳は，誤訳が多く，中には原文と似ても似つかなかったりとんでもない誤訳さえ散見されることに気付く．特にリアルタイムチャットの自動翻訳は（文脈を考慮しないため？）壊滅的に酷いことがよくある．

2026年現在の職業翻訳家のコンセンサスでは，最新AIの機械翻訳であれ，日英・英日翻訳の質と正確性はその分野のプロの翻訳家に遠く及ばない．特に，AIチャットボット翻訳は，前述の問題点と同様に「（いわば）嘘をつく」，具体的には原文の一部を丸ごと無視したり，時には原文にない文を作文したりする傾向がみられる（後者は“hallucination”の一種）．この傾向は，正確さを旨とする論文においては，致命的になり得る．

最善は，プロの翻訳家ないしは言語能力に長けた同業者に翻訳を依頼することなのは言を俟たない．その労を取らず敢えて機械翻訳を使うならば，その出力翻訳は細心の注意を払って吟味することが強く推奨される．そもそも専門家の仕事でも完璧とは限らず注意が必要なところ，まして機械翻訳はずっと劣るのだから．

実は，上記のAI翻訳の傾向は，2016年，機械翻訳サービスDeepLが登場した時にすでによく知られていた．DeepLは，深層学習の技術も含めたニューラル機械翻訳により，それまで最も著名な機械翻訳であったGoogle翻訳とは質的に「異なる」出力を生成することが特徴だった．出力された訳文が，自然言語としてGoogle翻訳の出力よりもはっきりとなめらかになり，自然に聞こえることが多かったものだ．いかにもネイティブの人間が言いそうな自然な翻訳出力文が衝撃的だった．

ただし，「自然に聞こえる」方が従来のGoogle翻訳に比べて常により優れた翻訳とは全く限らなかった．DeepLは，しばしば，翻訳原文の中の一部，それも単語レベルどころか複数文まで，全く翻訳せずに静かに無視する特徴があったのだ．対照的にGoogle翻訳では，翻訳エンジンが理解できない文章は理解できないまま，よく意味がわからない訳文が出力されていた．結果，ユーザーは，（出力言語を相当程度理解できるという前提で）出力訳文の一部が明らかにおかしければ，そこで明白に翻訳失敗している，と発見・同定できたものだ．DeepLのニューラル機械翻訳は，現代のAIチャットボットの原理と共通している．だからこそ，当時のGoogle翻訳と質的に異なる翻訳ができた一方，最新AIチャットボットに残る欠点も当時からすでに見られたと解釈していいだろう．

2026年現在も，Google翻訳は，DeepLや現代最新鋭AIチャットボットのように全面的に現代的（AIの）手法を採り入れることはせず，今も従来の機械翻訳手法とのハイブリッド方式を採用していると聞く．AIによる機械翻訳出力を検証する際，Google翻訳を使って比較することは一つの方法だろう．

3.3.2　翻訳が内包する問題点

翻訳には，それ自体に内包する問題点がいくつか存在する．以下，簡単に列挙する．

そもそも，自然言語は可換ではない．単語レベ

ルの例だと英語のsisterは日本語で姉または妹になる．あるいは，英語にある複数形の概念は日本語には希薄だ．結果，事前情報無しに「my sisters」を和訳するのは，姉か妹かあるいはその両方かがわからず，訳に困る．逆に英訳の場合，姉か妹かは明示しないのが普通だから，自然な訳では年齢情報が消える．日常会話ならば情報が少々消えても問題ないところ，科学論文の場合，これは望ましくないことがよくある．特に単数・複数形は数字に関わることもあり，その情報が翻訳によって消えるのは好ましくないことはあり得る．

また，英語と日本語では構文が全く違うことも，翻訳に大いに影響する．結果，翻訳の自由度は高くなり，それぞれ微妙にニュアンスが異なる複数の翻訳表現がそれなりに成立することが普通だ．

このいずれも，優れた翻訳者ならば，前後の文脈あるいは時には外部の独立の情報から，そして時には原文著者に質問することで，多くの場合，その状況に最も適切な訳文を提案できるものだ．AIでは簡単ではない．

さらに，日本語の構造として，最も大事なことは最後に持ってくる傾向がある．たとえば，「……が……をしない」など，否定の情報が最後にくる．その影響か，日本語母語の話者の場合，複数の文のつなぎ方や，さらにはパラグラフの論理構成においても，大事なことを最後にもってくるよう構成する傾向が散見される．英語の場合はこれは逆なので，そんな日本語を英訳するならば，パラグラフ構造まで修正した方がよい英語になることがある．とは言え，「訳し過ぎ」にもなりかねず，何が最善の英訳かは自明でない．

以上，一言でまとめれば，自然言語の翻訳は一般に可換ではないため，厳密さに拘る科学論文においては好ましくないことがよくある．そういう意味では，最初から英語なら英語で書いた方がよい，という主張に一理はあろうか．

4. 論文執筆および校閲で目立つチャットボットの特徴

2026年現在，科学論文の作文アシストにおいて，AIチャットボットの「さも本当のように堂々と嘘をつく」傾向（第2.3節）が最も端的に顕れるのは，

1）数字が不正確
2）参考文献が不正確．引用の文脈に無関係の文献を参照することがあるどころか，全く存在しない文献を引用することも多い．

の二点だろうか．前者の「不正確になる数字」とは，太陽半径のようなよく知られた数字ではなく，論文内または（マイナーな）引用文献内のどこかで出てくる数字であることが多い印象だ．後者は，ひどい場合はすべての参考文献が創作であった例も聞く―仮に気付かず盲信すれば，キャリアを失うスキャンダルにさえなりかねない．

結局，「統計的により多くのユーザーをより喜ばせる」（第2.3節）というAIチャットボットの動作原理により，チャットボットは事実上，「尤もらしく見える」体裁を整えることの優先度が最大になっていると言ってよさそうだ．AIチャットボットは，例えば「… was determined（Smith et al. 2020）」といった文が今までに出版された論文に典型的な英語表現であることは「知っている」から，そのような体裁の文を紡ぎ出そうとする．しかし，例えば後者における具体的な名前（Smith）や年（2020）の正確さは，気にしないようで，創作することがよくある．憶測すれば，数多くある過去の文献において格別な傾向はないからだろうか？　理由はともかく，AIチャットボット校閲は参考文献や数値に大いに無頓着なことは事実のようだ（より精度の高いチャットボットもあるかもしれないが）．

これらの欠点は，チャットボットに対して特定の工夫したプロンプトを送って回答に制約をあらかじめ与えておくことで，一定の改善が見られる

とは聞く．ただし，これらのプロンプトの「工夫」は，あくまで2026年現在，特定のAIクライアントに，特定の状況下でしか当てはまらない可能性があることはよく留意しておくべきだ．AIの進化によって変化するだけでなく，そのAIクライアントの運営企業の胸三寸でいつでも変わり得る．

上述の二点に加えて，チャットボットは，奇妙な論理構成，あるいは実質的に何も言っていないような論理構成を行うこともしばしばあるようだ．結局，チャットボットによる校閲や作文はあくまで，「部分的に見れば論文っぽい表現」つまり「世界の多くの研究者が頻繁に使用する表現」を紡ぎ出すことに特化したもの，くらいに考えておく方が無難かもしれない．

加えれば，その「研究者が頻繁に使用する」とチャットボットが判断した表現も，実は常によいとは限らない．論文の場合，英語の小説や契約書などと異なり，執筆者の大半は英語のネイティブスピーカーではないこともあり，ごく狭い特定の分野内に限って，英語としては本来少し不自然な特定の表現や単語が比較的頻繁にみられることがある．

筆者が依頼された英文校閲においても，耳障りな英語表現が気になって調べたら，それは実は著者の研究室の（日本人）先達の論文で（のみ）使われていた表現だったケースは複数回あった．つまりその英語表現は，英語非ネイティブスピーカーの先行研究者が何年か前に苦吟して紡ぎ出したものに過ぎず，端的にはほぼ誤用の英語表現だったわけだ（そして依頼者がそれを真似した）．

AIが過去文献の大量機械学習に基づいて「尤もらしい」表現を紡ぎ出す以上，こういった例外的かつ不適切なケースに影響されることはあり得るだろう．人間の校閲者である筆者は，そういう時，依頼者にその正確な意図を尋ねたうえで最適かつ自然な英語にするように校閲できる．AIは，一般に判断の根拠は示さない，ないしは示せない，こともあり，そのようなきめ細かい対応は期待しがたい．

5. チャットボットを使うのが不適切なケース―人間が担うべき役割

論文執筆における人間の最大の役割は何か？

まずは，論文の原型を執筆することだろう（執筆言語は問わない）．それは，八百万の事象からある未解決問題点を見つけ出して焦点を当て，それを（程度の差こそあれ）解決するまでのストーリー展開を構築し，それを言語化する作業にほかならない．そこに研究者のセンスと論理展開力が問われることは，今も昔も変わりない．

実際，筆者が請け負う論文校閲及び科学校閲において，通常，最も問題にするのは，次の3点だ：

1）論文中の論理展開における不備．端的には，論理の飛躍や欠如だ．
2）論文全体や各章内の論理構成とその整合性．論文の構成の原則は，イントロダクション章と結果章及び議論章との完全な呼応である．論文においては，結果章で述べる事実は，原則としてすべてイントロダクション章で何らかの導入がされているべきであり，逆にイントロダクション章で導入したすべての内容は，（イントロダクション章内で自己完結するものを除いて）結果章ないし議論章において一定の「答え合わせ」をすべきである．結果章と議論章とをどう切り分けるかの考察も必要だ．
3）定量化の不備．たとえば「大きい」という表現はそれ自体では単に「主観的な印象」の表明に過ぎない．科学論文では「AがBよりも統計的信頼度XX％でYY倍以上大きい」などと定量比較することが望ましい．関連して，有効数字の取り扱いの不備や論文内で数字が不一致のケースも散見される．

これらの査定は，投稿論文における査読者の仕事の重要な一部でもある．つまり，この3点に問題が

あれば，アクセプトが厳しくなることもあり得る．

そして，英語表現の校閲においては，科学論文である以上，

・著者の意図を正確に反映しているか
・定量的な表現になっているか

という点に最も気を配る．英語非ネイティブスピーカーが書いた英文の場合，著者の英語力次第で，著者が真に意図したことと，原稿の英文が意味することがずれていることも少なくない．最悪，（否定構文の誤用で）正反対の意味になることさえある．そんな時，内容をよく理解する人（たとえば共同研究者，あるいは科学的知識に優れた校閲者）ならば，前後の文脈から，英語表現のそのような重大な誤りにも気づいてくれるかもしれない．しかし，AIにはまだ荷が重いタスクと考えておいた方が無難だろう．悪いことに，そのような場合，「滑らかな文章に修正する」技術に長けたAIの場合，「著者の意図が正確に表されていなくてかつおそらくは不自然不明瞭な英文」を「清書」することで，原文よりも「英語的に滑らかで，それが故に著者の意図からは明白にずれた」英文に直してしまうことがままあろう．

また，後者の定量的表現については，そもそも原文中で定量的データが提示されていない場合，AIであれ人間校閲者であれ，定量的表現に修正することは不可能だ．筆者が校閲する場合，著者に「定量的表現が望ましい」とアドバイスすることがよくある．これもまた，AIにはまだ荷が重そうだ．

6. 論文執筆におけるAIチャットボットの将来と人間の役割

ここまで，AIよりも人間，すなわち論文著者自身あるいは（適切な能力ある）校閲者，の方が優れた点を列挙してきた．ただし，これは2026年現在の話であり，AIが急速に進歩している現在，従来は文句なしに人間の役割だったことのいくつかは，AIの助力も徐々に有効になっていく気配がある．そんな現在そして将来，人間が行うべき重要なポイントは何だろうか？

第3章でチャットボットが程度の差こそあれ有効活用できる事例を列挙した際，多くの項目は条件付きだった．一言で言えば，人間ユーザーがチャットボット出力の適切さを評価できることが条件だ．特に重要になるのが，読解力と論理把握能力・構成力だ．そしてこれは，少なくともAIシンギュラリティの日が訪れるまでは変わることはないと予想する．

現実に，大学教育の現場の教官から，「生成AIの出力をうのみにして論理的に一貫しない不自然なレポートを提出してくる学生が後を絶たない」という報告をSNSにて複数見かけた．そういった学生は，（単純に怠惰なのでなければ）AI出力を判断するだけの論理把握能力に欠けているわけだ．

もしそういう論理把握能力に欠ける学生が増えつつあるならば，それは社会として憂慮すべきことだろう．AIが急速発展している時代だからこそ，AIのアウトプットを判断できるだけの十分な論理把握能力と構成能力を備え，かつその能力を維持することが，高等教育を受けた人間やまして研究者には，今後今まで以上に不可欠になることと，筆者は予測する．少なくとも社会は，そういう能力ある人間を大量に必要とするに違いない．

また，現状，その能力を備え，問題解決にチャットボットを有効に使える人の場合も，時には立ち止まってそれだけでいいのかと自問する瞬間も必要かもしれないと感じる．なぜならば，問題解決をチャットボットに常に頼っていると，本質的な地力や応用力を損なう可能性がないかと恐れるからだ．

チャットボットをうまく使うスキルがあれば，過去遭遇した問題と同種の問題を解決するには十分だろう．しかし，将来，今までと質的に異なる問題が持ち上がった時，その人は対応できるだろうか？　入試の数学にたとえれば，公式を丸暗記するなどの最短ルートの試験対策で数々の演習問

題・過去問を短時間で解けるようになってきたとしても，教科書や参考書を熟読して系統的理解を深めない限りは，本質的な数学概念をいつまで経っても理解しないままになる恐れが残る．原理的には，大量の演習問題を解くことで，いずれ帰納的に本質的な数学概念を掴むことも可能は可能かもしれない．しかしそれこそまさにAIの機械学習のやり方にほかならない．そして，そのやり方では人間は，AIの桁違いの物量作戦に決して勝てず，惨敗するのは目にみえている．人間がAIに勝る点があるとするならば，系統的学習によって全体像を演繹的に掴む能力，そしてそこから派生する応用力ではないか？　あえて言えば，人間の矜持がそこにあるかもしれない．

チャットボット活用によって目先の問題が素早く解決できるようになったとしても，それに奢らず，自らの地力を高める不断の努力は昔も今も重要だと考える．チャットボットで「時短」を実現できたからこそ，本質的な勉強や自己研鑽により多くの時間と精力を割けるならば，それはチャットボット故の大いなる進歩になることだろう．

6.1　チャットボットの代わりに何を使うか？

物心ついた時からグーグルがあり，インターネットがない時代が想像できない若い世代の人が増えてきている．チャットボットも遠くない将来，同じことになるかもしれない．曰く，「チャットボットの問題点はわかるが，ではチャットボットを使わずにどうやって論文を書くというのか？」

その答えのひとつは，信頼できる（広い意味での）文献だ．教科書，時にはウェブサイトの記事や動画，そして言語の問題であれば（定評ある）辞書はAI時代の今も手放せない．信頼性という安心感は代え難い．

もうひとつは，ほかの人間だ．たとえば学生ならば，研究室の指導教官は多くの場合，最も身近で最も優れたメンターであろう．友人や同僚も外せない．そしてもちろん，科学文書編集や英文校閲に定評あるプロの校閲者に依頼するのは，最高に有効な手段になり得る．スマートフォン全盛の今の時代だからこそ，人との繋がりの価値は再発見されてよいと思う．

逆に，教育職の人々にとっては，AIにはできないか難しい「サービス」を提供できることが存在価値となる時代に移行しつつある現代かもしれない．たとえば，汎用のサービスを提供するAIに対して，精神面も含めて学生それぞれのニーズに応じた指導する能力は高い価値を生むことになることだろう．論文執筆ならば，クリアな論理展開を学生が見つけ出すお手伝いができれば，それは一つの理想になろう．AIは一つの（正答か誤答かを問わず）答を与える一方，教育者は答を導き出す力を育てる，と言い換えてもいい．

7.　外国語習得の意味

AIによる機械翻訳やはては機械同時通訳さえどんどん発展していっている現在，何千時間もかけて英語などの外国語習得をする価値がどれくらいあるか，という疑問を時に聞く．筆者は，3つの理由で，外国語（中でも英語）習得には今もそして将来も大いに価値がある，と確信している．一つは，研究者であれば現実に英語で論文を書かなければならない以上，AIを手助けとして使うことは可能ではあっても，最終的に判断する人間には相当の英語言語能力が不可欠であることだ（前述，第6章）．

また一つは，直接の会話，特に物理的に顔を突き合わせてのコミュニケーションにおいて，別の国の人と直接会話できることだ．友人として付き合いを深めるためには，現実的に共通言語を話せることは決定的に重要である．同時通訳は，仮に訳がすばらしくても，言語構造の違いにより，原理的にどうしても多かれ少なかれ時差ができる．たとえば日本語→英語ならば，日本語の文末の動詞が発話されるまで，その文の英語通訳が開始できないことは普通だ．そしてそもそも自然言語は可換ではない（第3.3.2項）ので，言語が違うと

すべて思ったように伝わるわけではない．加えて，言語はその国の文化や考え方とも関連するから，言語を知らずしては理解を深めるのが難しいことがある．人間は，ほかの人間と共生することに生き甲斐を見出す生物ならば，ボーダーレス時代の現代，より多くの人とコミュニケーションを取ることを可能にする外国語習得の価値はいや増していると思う．

もう一つは，異なる言語を習得することで，異なる思考形態を獲得できることにある．英語に長ければ，本質をまず最初に持ってくる，という考え方が自然にできるようになる．単純パターンの事象についてはAIやロボットが長足の進化を遂げている現代だからこそ，柔軟な思考こそ人間が誇ることになるのではないか？　少なくとも，多様な思考形態を獲得することは，その人の思考ひいては人生を豊かにすることになる，と主張する．

筆者は日本で生まれ育ち英国在住25年の今，日英バイリンガルであり，また過去にはほかに数ヵ国語勉強もしてきた．外国語を一つ獲得するごとに思考の地平が広がってきたことは実感としてある．むしろAI時代だからこそ，AIに負けぬように多様な思考を獲得するためにも，外国語を積極的に習得することを推奨したい．

8.　まとめ

本稿では，研究者の言語活動における生成AI型チャットボット活用方法について考察した．AIチャットボットは機械学習をベースとするから，学習機会に事欠かないビジネスレターなどでは，論文の査読審査員との折衝も含めて，特に有効になる．現在のAI，そしておそらく近未来のAIも，「客観的に正しい回答」よりも「統計的により多くのユーザーをより喜ばせる」を目標にしている．その結果，AIチャットボットは，さも本当のように堂々と嘘をつく傾向があり，特に論文執筆に関しては，数字と参考文献の嘘や誤りが目立つ傾向がある．また日英翻訳では，原文の一部を無視したり，逆に文をゼロから創作したりする傾向がある．これらリスクを十分に意識しながら使う限りは，論文執筆（校閲，翻訳）においても，AIチャットボットは有用な場面が少なくない．本稿ではその具体例も挙げて議論した．ただし，条件として，著者自身にチャットボット出力の適切さを判断できるだけの言語能力と論理把握力が欠かせない．結局，AIは上手に利用すれば救世主並の時短になり，何も考えずに使えばキャリアを破壊するレベルの恥をかかせられる悪魔にもなる，と言える．いずれにせよ，急速に発展するAIと付き合わざるを得ない現在ならびに未来において，人間の言語能力，論理構成力，系統立てた知識，判断能力はますます重要になる，と考える．

Is the AI Chatbot a Messiah or a Demon for Writing English Academic Papers?

Masaaki "Masa" Sakano
Wise Babel Ltd, Scotland, https://wisebabel.com

Abstract: Generative AI chatbots have made a breakthrough advancement in broad fields in recent years. Here I lay out their general characteristics and give some proposals about how scientists, particularly non-native English speakers, should use or avoid them to help writing papers in English, weighing their benefits and risks, based on my years long professional experience of scientific editing and English proofreading. Although the fact that AIs are optimised to satisfy the largest number of users, sometimes in sacrifice of veracity and accuracy, is disadvantageous in the field of science, it does have very significant benefits in the field, if used carefully. I argue the human user's ability to evaluate the AI's outputs is more crucial than ever.