



日本公認会計士協会
テクノロジー委員会
未来の監査専門委員会



生成AIの可能性と会計監査への活用（前編）

～山田誠二国立情報学研究所教授と未来の監査専門委員会との意見交換～

1 はじめに

近年、様々な業界でAI(人工知能)の活用が進み、監査業務においても、AIの開発及び利用の事例が出始めており、今後さらに広がりを見せようとしている。とりわけ、2022年11月に公開された「ChatGPT」をはじめとした生成AIは、AIに関する専門的な知識がないユーザーにとっても、自然言語を利用したQ&Aシステムのように利用することができるため、監査チームにとってAIを利用する敷居が下がっている。生成AIをどのように使えば監査をより効率化できるのか、

監査事務所レベル及び個々の監査チームのレベルでも、日々試行錯誤されているのではないかとと思われる。

日本公認会計士協会テクノロジー委員会(未来の監査専門委員会)では、2024年8月13日に、テクノロジー委員会研究文書第11号「監査におけるAIの利用に関する研究文書」を公表したが、当研究文書は、生成AIに限らないAI全般を取り扱うものであった。そこで、今回は生成AIによりフォーカスし、生成AIの特質、他の業界での利用事例や、生成AIとの関わり方などについて、国立情報学研究所の山田誠二教授と意見交換を行った。

山田誠二：

国立情報学研究所教授、総合研究大学院大学教授(2025年6月23日時点)
(2025年9月より、神奈川大学情報学部システム数理学科教授)

HAI(ヒューマンエージェントインタラクション)、IIS(知的インタラクティブシステム)の、2つの研究分野を世界に先駆けて展開し、人とコンピュータシステム、マシンとの協調を目指して、人工知能や計算知能の手法を駆使したインタラクションデザインの工学的設計論を研究している。

(テクノロジー委員会 出席者)

※役職は意見交換を行った2025年6月23日時点のもの

伊藤公一：専門委員長

有久衛、井上敏、佐々木崇、清水希理子、須崎公介、多田野靖記：専門委員

小林尚明：常務理事(テクノロジー担当)

紫垣昌利：テクノロジー委員会委員長

佐藤重義：テクノロジー委員会副委員長

2 意見交換の概要

① 生成AIの特質

(山田誠二 国立情報学研究所教授)

皆さんも人前で挨拶を述べることがあると思います。あらかじめ挨拶文を準備する際、定型的な表現をネット等から拾ってこれを加工して作ることも多く、全くのゼロから作る方は多くないと思います。挨拶文のような定型文の生成は、生成AIの最も得意とするところ。生成AIにプロンプト(指示)を与えると、プロンプトで教えていない内容も自動生成されることがあります。これは、生成AIのLLM(大規模言語モデル)が挨拶の定型文などのようなインターネット上にある膨大な文章をあらかじめ学習しており、トランスフォーマーというモデルによって文脈を理解し、関連する情報を補完しているためです。LLMは、プロンプトとして与えられたN個の単語列から、次に続くN+1個目の単語を予測することで文章を生成しています。例えば、「I have an」と入力すると、次に「apple」がくる確率が最も高いといった形で、確率分布に基づいて次の単語を予測します。Nが長ければ長いほど予測の精度が上がります。ただし、N個の文字を均等に扱っているわけではなく、文脈を考慮して重み付けをした上で、次の単語を予測しています。

この結果、ネット上に公表されている当たり障りのない表現が多く生成されます。挨拶文等は、この当たり障りのない内容を事前学習して定型化しやすく、生成AIは、このような文章であればものすごいスピードで自然な文章を生成できます。また、長い文章を要約することも得意です。

一方で、クリエイティブな内容はネット上に多く存在していないため、そのような文章の生成は不得意です。こういった生

成AIの得手不得手をよく考えた上で使うことが重要となります。

また、生成AIはURLを半分に切って入れ替えてくっつけたり、人の苗字と名前を取り違えるという誤りを犯すことがあります。そのようなことをすると意味のないものになるという人間的な常識を生成AIは学習できないためであり、ハルシネーション¹といわれるLLMの失敗例の多くはここに起因しています。

生成AIの特徴としては、入出力が自然言語を用いたQ&Aシステムであることや、多言語に対応していることが挙げられ、プロンプトエンジニアリング²を通じてエンドユーザー自身がシステムに、適切で必要十分な指示を与える必要があります。生成AIに1のプロンプトを入れて10の回答が返ってくる場合、残りの9の部分とうまく引き出すように1を入れる必要があります。人間側のプロンプトエンジニアリングのスキルは重要となります。

② 人間と生成AIとの協働(人間・AI協調意思決定)

上司と部下が共に仕事をする際、上司は部下の能力を意識的又は無意識に推し量り、上司と部下の合計パフォーマンスが最大になるように、上司は部下に業務指示を与え、部下の業務結果をどの程度レビューするかを決めると思います。生成AIの場合も、生成AIを1人の部下と見立てた上で、生成AIの能力を人間が推定し、人間が全て行うか生成AIに任せるかを人間が意思決定することが重要です。理論的には、同じタスクを人間とAIの双方に実施させ、タスクの成功確率を出すということを繰り返し、その結果を踏まえて成功の可能性の高い方を選択することとなります(人間・AI協調意思決定)。

また、上司は部下にタスクの指示を出す際、指示内容に曖昧な部分があっても、部下が常識に照らして適宜相応な解

釈をしてくれると思っており、実際そうなることも多いですが、AIには人間的な常識がなく、思わぬ間違いもよく起こるので、それを避けるように人間側でプロンプトを与えるということも重要になります。一方で、生成AIに文章を生成してもらうために人間が事前準備や事後の手直しに時間をかけすぎると、結局人間が文章を作成した方が早いということになりかねないため留意が必要です。

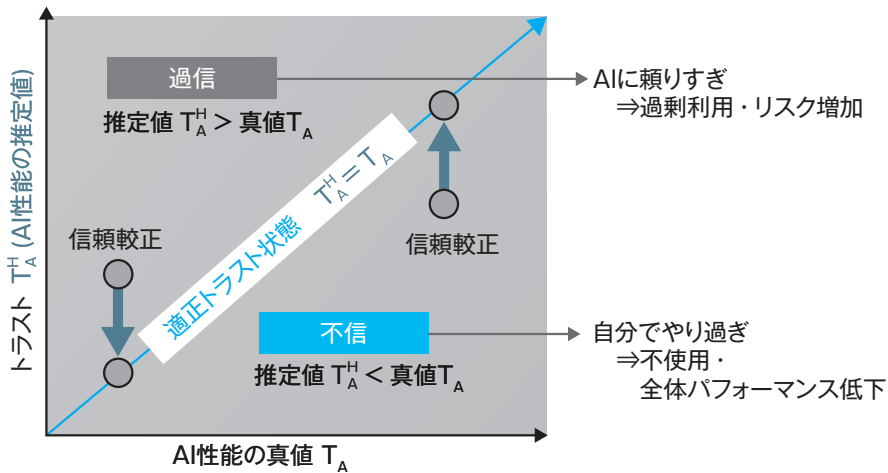
③ AIの信頼(トラスト)と信頼較正(医療AIの例)

あるタスクをAIが実施した場合の成功確率に関する人間の推定値を信頼(トラスト)と呼びます。トラストは、認知科学・社会心理学においては、その人の能力や実績に基づく認知的信頼(cognitive trust)のほか、人間関係や社会規範などに基づく感情的信頼(emotional trust)がありますが、AIの場合、感情的信頼の部分は無視できるので、認知的信頼のみで性能を評価します。

次頁の【図1】は、縦軸にAI性能の推定値、横軸にAI性能の真値をとったものです。推定値と真値が一致する45度線の上にトラストがある状態が、人間がAIの性能を正しく評価できている最適信頼の状態であり、人間とAIを合わせたパフォーマンスの最大化につながることは自明ですが、この45度線より上にある場合、人間がAIを「過信」しており、また45度線より下にある場合、「不信」の状態にあるといえます。「過信」の場合には、AIを過剰に利用することでリスクが高まる状態であり、逆に「不信」の場合には、AIを過小評価し使用しないことで全体のパフォーマンスが低下している状態になります。ここで、「過信」又は「不信」の状況から、45度線の上にトラストを補正することを「信頼較正」といいます。

医療業界では、「信頼較正AI」というAI

【図1】過信・不信・信頼校正



がトラストの状況を判定し、人間にトラストの見直しを行わせるという実証実験が行われています。例えば、肺のX線写真を読影して診断を下すタスクを大量に抱えている医師が、読影作業の一部をAIに任せる場合があります。その際、高齢者のX線写真は白く濁る傾向が強く、AIが適切な判断を下せない場合が多くあります。そのような状況で高齢者の読影をAIに任せると、信頼校正AIが過信を検知し「AIに任せる判断を見直した方がよい」と、判断見直しの促し(校正キュー)を発するわけです。医師は信頼校正AIの促しを受け入れ、高齢者のX線写真の読影はAIに任せず医師が自ら行うように注意する一方、クリアなX線写真が撮影できる若い方のX線写真の読影はAIに積極的に任せるといった判断を行えるようになります。このように、人間とAIがインタラクションを通じて、AI性能を正當に評価することにつながります。

④ 意見交換

AI利用時の専門家の判断

(伊藤) 紹介いただいたX線写真の読影の事例ですが、信頼校正AIの促しを受けて、若年者だからAIに任せて大丈夫、高齢者は医師自身が行おうという

医師の判断は、最終的には個々の医師の判断なのではないでしょうか。信頼校正AIの促しが正しいか否かについて、医師がどのように判断しているか気になりました。

(山田) 現在では、最終的に医師の判断になります。高齢者の方の読影の難易度が高いということを医師が経験値・暗黙知として持っていることも重要です。このような知見は、医学の授業等では教えられていないし、使用しているX線写真撮影の機材の性能や解像度によっても状況は異なります。したがって、医師同士の交流は非常に重要であるといえます。

AI利用時の顧客への説明義務

(多田野) AIに読影・判断を任せて誤診が生じた場合、最終的に責任をとるのはAIの診断結果を採用した医師自身であることは理解しましたが、AIに診断を任せることについて、患者への説明義務はありますか。また、生成AIによる診断結果を医師がどのように取り扱わなければならないか、具体的なガイドラインはありますか。

(山田) 患者にAIの利用形態及び利用方法を伝える義務の有無についての議

論はまだ始まっていないと理解しています。最終的な診断を下したのは医師であるとしても、AIの診断結果に影響を受けていることは間違いありません。現在では、医師がAIの診断結果をどのように利用するか具体的なガイドラインはまだありません。この点、AIの技術的発達に伴い、AIによる診断結果を医師が覆した結果として誤診が生じた場合、医師により重い責任を負わせるというガイドラインが議論されていますが、そうすると、医師はAIの診断結果を覆すことをしなくなり、AIの過信につながるバイアスが生じることも考えられます。いずれにしても重要な最終判断を行う主体は、責任を負う主体である必要があります。

AIリテラシーの教育方法

(伊藤) AIの診断結果に対する判断が医師により異なると思いますが、医師に対してAIの技術やリテラシーを身に付けるための教育の動きは業界全体としてありますか。

(山田) 医療業界でもAI教育に関する問題意識はあると思いますが、まだ進んでいないと認識しています。この点は医療業界に限らないのではないかと思います。一方で、生成AIへ質問を入力し回答を得る作業は、分からないことを人間に聞く場合とどこが違うのかという議論や、生成AIが行っていることは人間がウェブで検索して出てきた結果をまとめる作業とあまり変わらないのではないかという議論もあります。

AIの出力結果への依拠

(井上) 監査の現場では、AIを使って証憑の突合や開示書類の検証を行うことが始まっています。作業を完全にAIによる手続実施結果に依拠したいので

すが、現時点ではAIが処理結果を誤ることや、重要な点を看過するリスクもあるので、人間が再確認しなければならない場合が多いです。人間が結局確認するのであればあまり効率化しないという意見もあり、もしかしたら「不信」の状況に陥っている可能性もあると思いますが、どのように考えればAIの手続実施結果に依拠できるようになりますか。

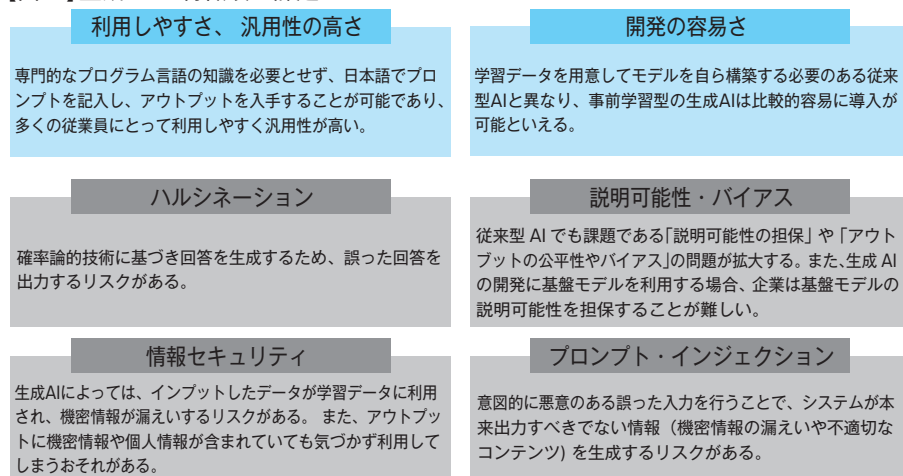
(山田) AIの性能(間違える・看過するリスク)の見積りは、答えが分かっている問題を多く解かせて、その正答率を出すことで推定できます(最尤推定)。自動車の自動運転など、AIの判断が誤れば即交通事故につながるような取り返しのつかない状況であればAIの性能は相当に高くせざるを得ませんが、個々の監査手続で利用する場合、人間による手続実施結果の正答率も100%ではないとすれば、人間並の性能が出ればAI利用による効率化の余地があるのではないのでしょうか。どの程度の性能のAIの利用結果が社会的に許容されるか、各業界においてガイドラインを発出する等によりコンセンサスを形成することも重要であると考えます。

会計上の見積りにおける経営者の主張をAIによって監査できるか

(小林) 重要な監査領域として「会計上の見積りの監査」があります。この対応には、経営者による仮定の背景にある経営理念や経営者の考え方がかわってくることも結構多く、これをAIが監査するのは難しいと思っています。

(山田) 経営理念や経営者の考え方の全てを文書化・言語化することが可能であれば、AIでも監査手続を行う余地はありますが、言語化できない部分につ

【図2】生成AIの特徴及び課題



いては、LLM等で対処するのは、小林さんがおっしゃるとおり難しいのではないのでしょうか。経営者の考え方は、話し方や表情、動作に顕れることもあると思いますが、人間であったとしても、オンラインミーティングではこういった情報を的確に捉えることは難しい場合もあると考えます。

監査における生成AI利用の課題

(伊藤) 監査における生成AIの特徴及び課題として、【図2】に記載の内容があると専門委員会では考えています。こちらについて意見はありますか。

(山田) 監査業務では被監査会社の秘密情報を取り扱っており、漏えいすると監査の信頼性に影響するため、情報セキュリティは特に重要だと思いました。クラウド上ではなくローカル環境で動作するLLMの活用も1つの防御策として考えられます。

説明可能性の点ですが、軍事のオペレーションでは、飛来物が敵かどうかを瞬時に判断して迎撃するという場合がありますが、誤って味方や民間の航空機を迎撃してしまった場合に、なぜ間違えたかという説明が必要となります。私は監査の専門家ではありません

が、この点、監査業務では、AIの回答の説明可能性を担保することは不可欠でしょうか。説明の部分人間がカバーするということはあるのでしょうか。

(伊藤) AIの判断に基づいて、被監査会社に調査を依頼し、場合によっては修正を依頼することも想定されますが、依頼の根拠が分からないと被監査会社に納得してもらえないため、説明可能性は重要であると考えています。説明の部分人間がカバーするケースもありますが、根拠が全くブラックボックスで人間には想像がつかない場合には、難しい対応になると考えています。

生成AIの精度の限界

(小林) 監査は社会的なインフラであるため、信頼性の担保は非常に重要だと思っていますが、生成AIを利用しているとハルシネーションが生じて、信頼性を担保できないと感じることも多くあります。将来的に、ハルシネーションが解消されて生成AIの結果をより信頼して依拠できるようになるのか、山田教授のお考えをお伺いしたいです。

(山田) 大量のデータを学習させることで



物を適切にレビューしながら、積極的に使っていくことが望ましいと理解しました。

山田教授のお話の中では、生成AIを部下に見立てて、タスクを依頼するかどうかを検討するという例が印象的でした。生成AIは人間とは異なる特徴を持つ新たな監査チームの一員であり、正しく使いこなせば重要な戦力にもなり得ます。

未来の監査専門委員会では、生成AIを監査においてどのように活用するのがよいか、今後も引き続き、調査研究を進めてまいります。

後編では、山田教授との意見交換を踏まえ、未来の監査専門委員会において、公認会計士が生成AIを具体的に活用する際のポイントや注意点、また生成AI活用に向けての人材育成などについてディスカッションした結果を報告したいと思います。

<注>

- 1 ハルシネーションとは、生成AIが、事実に基づかない情報を生成する現象のことをいいます。
- 2 プロンプトエンジニアリングとは、生成AIに対して適切な指示を与えることで、適切な回答を引き出すための技術をいいます。

生成AIの性能が向上していますが、生成AIのアルゴリズム自体の性能が今後飛躍的に向上する可能性は低いと考えています。利用の領域や条件を絞り込めば100%に近い性能が出ることもあります。絶対には間違えないわけはありません。90%や85%の性能のものをいかに使いこなすかというところが重要ではないでしょうか。人間の採用活動でも、100%の能力を有する方のみを採用しようとしても、うまくいかないのと同じです。

AGIの実現可能性

(紫垣) 近年、人間のような汎用的な知能を持つAGI(Artificial General Intelligence：汎用人工知能)が将来的に開発されるという記事を目にしますが、AIの得意な範囲が少しずつ広がった結果、AGIに到達するというイメージでとらえればよいでしょうか。

(山田) 様々な意見があるところですが、AGI研究の方法論をみると、人間の脳をベースに行われているようです。ただし、人間の脳の学習とコンピュータの中で行う機械学習は全く異なるもの

です。人間の脳にはスーパーコンピュータの何万倍もの並列計算機が入っているような状況であり、そのようなプログラミングはできないこと、また人間の脳の働きをモデルとしたニューラルネットワークの何千万倍もの関数が脳の中にはあるといわれています。生物が何億年という時間をかけて進化し、一団体が膨大な環境との相互作用を通じて学習してきたプロセスを、現在のコンピュータ技術で再現することは、計算能力の観点からも、学習メカニズムの違いからも、非常に困難であると考えます。

3 おわりに

(伊藤) 本日はありがとうございました。AIの技術が進化し、一部の作業や判断について、AIへの代替が進んだ場合でも、重要な監査手続や判断の実施をAIが代替することは難しいと改めて感じました。人間がまず生成AIの特徴や得手不得手を理解し、生成AIの能力を正しく評価して、100%の精度は期待できないことを認識した上で、成果

*法定監査従事者の必須研修科目「監査の品質及び不正リスク対応」研修教材	
教材コード	J 0 3 0 5 8 3
研修コード	3 1 9 3
履修単位	0.5 単位