

# 中央大学 会計人会 会報

発行所 中央大学会計人会

〒115-0045 東京都北区赤羽1丁目52番1-501号田口ビル

税理士法人 石亀総合会計事務所

<http://chudai-kaikeijin.jp>

発行人 会長 石亀 邦 俊



## 生成AIの最近の動向と 会計業務へのインパクト

中央大学理工学部物理学科 田口 善弘

### 生成AIとは？

昨年（2022年）の11月に一般に公開されたChatGPTは世間の耳目を集め、大きな話題となった。同時に、あたかも人間とみまごうようなやりとりをやってのけることから、人間の職を奪うのではという懸念が惹き起されるに至った。本当にChatGPTが人間の職を奪うほどの能力を持っているか否かということについてはまずは浅くだけでもChatGPTがどういうものでどうやってできたのかを知る必要がある。

ChatGPTは大規模言語モデル（Large Language Model、以下、LLM）の一種であり、LLMはいわゆる生成AIと呼ばれるものの一種である。生成AIはその名の通り、何かを生成するAIである。LLMは言語（文章）を生成するが、現在、多くの生成AIが開発されているがLLMはその一種に過ぎない。現在いわゆる生成AIが作り出せるものとしては、言語のほかに画像、動画、音楽、3Dモデル、などと多岐にわたっている。異なった言語の間をつなぐいわゆる機械翻訳も生成AIの一種であると言ってよいであろう。

これら生成AIの特徴はなんと言っても新しい物を作り出す能力にある。ChatGPTのリリースに数か月先んじて、既に画像生成AIが大きなブームとなっていた。この画像生成AIの特徴は言葉で「○○を描いてください」と命令すると、いい感じに絵を描いてくれる。まだ試したことが無い方は是非、試してからこの原稿の先を呼んでほしいのだが、LLMを内包した検索エンジンの1つであるBingにはBing Chat [1] という機能がある。スマホかインターネットにつながっているパソコンでBing Chatにアクセスして「○○を描け」と入力して欲しい。おそらく皆さんはそのクオリティの高さに衝撃を受けるだろう。絶対にどっかの誰かが描いた絵にしかみえないはずだ。だが、それは生成AIの制作物であり、決してどこかにあるものを取ってきたものではないのだ。

当時はまだLLMが現在の様に誰でも使えるような環境には無かったため、「○○を描け」と入力すればいい、というほど簡単ではなかったものの、非常に簡単に人間だったら相当な練習を積み重ねなくては描けない絵を生成AIが簡単に描けるようになった

ことで大きな議論が巻き起こった。誰が見ても「これは人間のイラストレーターが失業するな」と思うようなクオリティだったからだ。議論が巻き起こった理由はこの生成AIがどうやって開発されたかに大きく関わっていた。この生成AIは無から絵を作り出しているわけではなく、インターネット上に誰でもアクセスできるような形で公開されているデジタル画像をダウンロードしてきて生成AIに学ばせて初めてできるものだったからだ。

元の画像と同じものを出力しているわけでは無いから、直接著作権に違反するわけでもなく、どこの誰の画像を読み込んだかも明確では無いため、元の画像を作った人間が著作権の侵害を申し立てることもできない。なにより問題だったのは、元の画像を描く労力に比べて生成AIが出力する画像の方が圧倒的に出力する速度が速かったことだ。これでは人間のイラストレーターに頼む人間などいなくなってしまう。

実はこの「人間の作ったものを生成AIが学んでしまって元を作った人間を失業させる」という問題はこの画像生成AI以前にすでに起きていたのだが、やや小規模だったので大きな問題として取り上げられにくかった。それは機械翻訳の分野だった。皆さんはインターネット上でアクセスできる自動翻訳を使ったことがあるだろうか？ ちょっと前まで、機械翻訳は全く使い物にならないほどひどい物だったが、しばらく前に急に性能が上がって使えるようになった。これも例えば大量の英語-日本語翻訳文を生成AIが学んだ結果可能になったものなのだ。

これら生成AIが今後どのような発展を遂げ、どんなことができるようになるかを語る前に生成AIができた歴史を簡単に振り返っておこう。

## 生成AIの歴史

生成AIのきっかけとなった「事件」はいくつかあった。生成AIのAIとはArtificial Intelligence、つまり人工知能である。人工知能の研究は長い歴史があるが、最初は人間の知能をプログラミングで置き換えて人間と同じことをさせようという試みが主流だった。しかし、この試みは遅々として進まず、有体に言って暗礁に乗り上げているような有様だった。なぜ、これが難しいことだったかはここでは述べない。このような試みと並行して全く別の方向からの試みが誕生した。それは「機械学習」と呼ばれる分野である。人間の知能を明示的にプログラムに書き下そうという王道的なAIの研究とは異なり、機械学習の方法ではいろいろな方法でコンピュータに学習する能力を与えて勝手に学習させる。例えば、犬の写真と猫の写真をたくさん用意して、「こっちは犬、こっちは猫」という分類と共に機械学習に勉強させると、初めて見る写真でも「犬か猫か」と答えられるようになる、というものだ。この場合、機械学習が勝手に学習するので往々にして人間は機械学習が写真をどう解釈して答えを出しているか解らないのだが、一定程度うまく行くことが解っていた。

この機械学習の一種にニューラルネットワークという物があり、これは人間の脳のニューロンを模したものだ。だが、このニューラルネットワークは他の機械学習に比べて性能が上がらず、20世紀末には一旦忘れ去られた。

だが、21世紀になってからしばらくしてこの状況は一変した。役に立たないと思われていたニューラルネットワークをしぶとく研究していた研究者が深層学習 (Deep Learning,以下DL) というニューラルネットワーク

ワークの改善版が非常に高性能を出すことに気付いたのだ。このDLとニューラルネットワークの違いはたった2つしかない。ネットワークの層が数層から百層以上に拡大したこと、学習するデータの量が膨大になったことだ。20世紀末にこれが出来なかった理由は単純で、百層のニューラルネットワークを学習するだけの高速なコンピュータが無かったこと、また、インターネットが普及していなかったので学習に必要な膨大なデータがそもそもなかったこと、そして、仮にそれがあってもそのデータを処理できるだけの大規模なメモリーがなかったこと、だ。この事実気付いたDLの考案者であるヒントン教授は、自らのDLを画像判別（画像を見て何の画像かを充てる）コンテストに参加し、ぶっちぎりの好成绩を納めた。

このDLはほどなく、画像判別だけではなく、他の機械学習にも有効だということが分かった。その1つはゲームである。将棋、囲碁、チェスなどのゲームで人間に勝てるコンピュータを作ろうという試みは長い間なされてきたがAIの研究が遅々として進まないのと歩を合わせるように遅々としてすすまず、この3つの中でもっとも簡単と言われるチェスにおいて、20世紀末にやっとコンピュータが人間に勝つという有様で、取った駒をつかうことができるためにチェスよりはるかに複雑な将棋や、盤面が将棋やチェスに比べてずっと大きい囲碁は、いつ人間に勝てるのかようとして知れなかった。だが、DLが「今の盤面に対して最良の刺しては何か？」という一種の判別問題に置き換えて解くことで囲碁や将棋でプロ棋士に勝てるようになった。いまでは人間のプロよりコンピュータが強いことを疑うものはいないまでになった。

そして、画像とゲームに続いて、言語も

DLで制覇されるようになった。当初、言語は画像やゲームよりはるかに難しく、例えば、大学入試の国語の問題を解くようなことは当分できないと思われていた。それは何より、機械学習が学べるほどの大量の文章題が無かったからだ。

この時点で生成AIにつながる大きな革命が起きた。ある研究者が大学入試の問題を解くDLを学習させるには必ずしも入試問題でなくてもよく、普通の文章でいいことに気づいたのだ。その辺にある文章をただ持ってきて、文章を部分的に隠し、隠したところを当てさせるという簡単な問題をたくさん解くだけで、最後にちょっとだけ文章題をやらせればいい成績を取ることができる、と解ったのだ。

これは最終的な目的（入試問題）と学習しているもの（普通の日本語）が同じでないため、この一般的な学習を「基盤モデル」と呼んで特別に扱うようになった。最初は言語だけだった基盤モデルは、その後、何でも使えることが解った。画像だろうが、音楽だろうが、ともかくDLでたくさん学習させておけば、それで絵を描かせたり、音楽を作曲させたりすることができることが解ったのだ。

そこでこれらを生成AIと呼ぶことにした。LLMは生成AIの一種であり、言語を作り出す生成AIとして誕生した。実は話題のChatGPTは言語の基盤モデルに、入試問題ではなく会話を学習させただけのものである。人間の問いかけに対して、人間らしい答えができるように人間が付き合っただけ人間らしいふるまいを教え込むとなぜかあたかも人間が答えているかのような応答ができるようになったのだ。

## 生成AIのからくり

現在はまだ、この生成AIがなぜここまでの高性能を発揮して音楽や動画や画像を生成できるのかはよく解っていないが、ここではいま考えられていることを簡単に説明しておこう。この節で書くことはやや難しいのでし解らなければ読み飛ばして欲しい。

生成AIがこのようなこと、つまり、学んでいないことを「生成」できるようになった鍵は、「意味のある文字列は稀である」ということに尽きる。例えば100字の文章を考えよう。単に「100字の文章」というだけなら、この世には無限ともいえるほどの文章がありうるだろう。日本では毎年7万冊程度の書籍が刊行されている。一冊の本に10万字の文章があるとすると、実に70億文字の文章が刊行されていることなる。これを100字ずつにくぎったところで全部意味がある文章にはならないだろうが、仮にそうすると70億を100で割って7000万個の、100字の文章が生み出されている。これはいったい「あり得る100字の文字列」のうちのどれくらいの割合だろうか？ここでは文字は50音のひらがなに限定とする。それでもたった2文字の文章でもありうる数は「ああ」から始まって「んん」までで $50 \times 50 = 2500$ 通りになる。3文字ならさらにその50倍で12万5千通りである。計算してみるとたった6文字の文章で軽く70億を越えてしまう。ありえる100文字の文章の数は膨大で、8の後にゼロが169個つく数になる。一億は1の次にゼロが8つつく数なので、169割る8は大体21であることから、平仮名だけからなる100文字のあり得るパターンは数は一億を21回かけたくらいの気が遠くなるような膨大な数である。つまり、まったくランダムに文字を並べたら、意味がある文章ができ

る可能性はほぼゼロだ、ということになる。

さて、ここで100字の文字列に「番地」をつけることにする。通常の番地は地表面についているので2次元で、例えば、緯度と経度だけで地表面を表現できる。だが、ここではもっとずっと高次元の次元が高い、何万次元もの空間を考えて、その膨大な数の100文字の文字列を配置する。機械学習の研究でわかったことはうまくやると「意味のある文章」だけをごく一部に集めて、他の意味のない文章はそこにはないように「座標」を作れる、ということだ。地球上にすべての100文字の文章をちりばめたとき、意味のある文章は東京都に集中している、みたいな状況を考えてほしい。

さて、ここからが肝なのだが、どんなたくさん文章を集めてきても絶対に文章と文章の間には隙間ができる。では、意味がある文章が集まっている隙間の一点を差し機械学習に「この隙間を埋める文章を作れ」と命じたらどうだろう？それはほぼ確実に意味がある文章になるだろう。生成AIであるLLMがこの世にない文章を作れる秘密はこの「意味のある文章だけをまとめるような座標化が可能だということ」をみいだしたことにある。この結果「今日は雨です」という文章と「今日は晴れです」という文章の隙間の一点を指し示し、機械学習に「この隙間を埋める分を作れ」と命じると、「今日は曇りです」という一度も学んでいない文章を作る。機械学習は文章の意味など全く理解してないが、意味ある文章のそばには意味がある文章があることは知っているから、意味も解らないのに意味がある文章を作るという離れ業ができるようになった。

同じことは画像や音楽でも簡単にできる。我々が目にしているデジタル画像は拡大してみれば色がついた点の集合に過ぎない。もし、この点に「勝手な色」をランダムに

置いたら人間にとって意味がある画像になる可能性はほぼゼロである。文字の時と同じように、意味がある画像が集まっている場所の空き地を指し示し、「ここを埋める画像を作れ」と機械学習に命じれば、機械学習はそれっぽい画像を無限に作るだろう。

音楽もデジタルであればいろいろな高さの音の連続にすぎないので、文字や画像と同じように考えれば、「意味のある音の列 = 音楽」は稀にしか存在せず、音楽を集めた場所を作って置き、その隙間を指し示せば、音楽を奏でる機械学習が出来上がる。これはデジタルデータで表現できるすべてのものに応用可能だということは容易に想像できるだろう。

言葉で指示すれば絵を描く機械学習を作りたければ、絵とそれを説明する文字の組をたくさん与えて「座標」を作らせればいい。そのような機械学習は与えられた文字列に一番近い絵を生成する機能を持つようになるだろう。これがまるで魔法の様に言葉で何かを作り出す生成AIのカラクリである。

LLMに「絵を描け」という命令が出されたら、裏側でこっそり「文字から絵を描く生成AI」を呼び出してできた絵を見せれば、人間には「言葉で指示すると絵を描くAIができた」と感じられるだろう。これこそが冒頭で紹介した「言葉で指示すると絵を描くAI」のカラクリである。これらについてより詳しいことはわたしの中大の学術講演会の記録 [2] のうち「AIの過去・現在・未来」と題する講演（特に2023年度のもの、2019年度のもの）を参照して欲しい。講演ビデオと使用スライドが公開されている。

## 会計業務へのインパクト

さて、最後に会計業務へのインパクトについて語りたい。僕は専門ではないのだがLLMを用いて会計士の試験をパスできるかどうか試した研究は存在する [3]。LLMは日進月歩であるのでこの結果をずっと信じるのは問題があるがあくまで参考程度であれば十分に役に立つ内容である。この原稿の公開日は2023年10月であり、時期的には2023年度の上半期のChatGPTの性能で評価されたと思ってよい。[3] は非常によくかけているが如何せん人工知能学会での研究報告であるため専門用語が頻出し、予備知識がない場合は難しいので以下、かみ砕いた解説を行う。正確さに多少劣る部分、とくに公認会計士試験についての知識不足から来る過誤についてはご容赦願いたい。

まず、現在の試験は4肢6択といって4つの選択肢から正解2つを選ぶ形式になっている。しかし、これをLLMにやらせるのは現状では困難なので、4つの選択肢が独立に正解か不正解かを当てさせる問題として学習させた。これでは正しい2つを選ぶ確率は解らないのであるが、これにはいくつか理由がある。まず、前節の説明でわかるように学習にはある程度の数が要る。4肢6択を4つの正誤問題とすることで学べる例題が単純に4倍になり、機械学習の精度があがる。また、現在のLLMでは、これまた前節の説明からわかるように、選択肢の提示順が異なっているが選択肢の中身が同じであるものを同じ選択肢群であると認識するのが困難である。あくまで文章を文字の羅列としてとらえているだけで意味を理解しているわけでは全くないので、LLMに取っては同じ選択肢が違う順番で提示されたら別の情報になってしまう。このためLLMからすれば違う入力に対して同じ出力

をしなくてはならないという高度なことを要求されてしまうので学習が難しくなる。以下の記述はこのような簡単化を行ったうえででの評価であることを理解されたい。同時にこのような簡単化が必要だということがLLMの限界を表しているのとすることも出来よう。

また、単純に選択肢を提示するだけではなくいろいろ条件を付加している。LLMは「専門家だと思って答えよ」と言われた場合と言われない場合では答えの精度が変わることが知られている。これはある意味で当たり前でChatGPTが学んでいるのは「人間らしい応答」なので特に設定もないのに専門家の様に答えたら却って不自然なので必ずしも正しい答えをすることが「人間らしい応答」としてベストではないからだ。今回のテストでは専門家として答えるような条件設定を問いかけの中に入れてることをまず注記したい。

また今回は予め専門家として追加学習をしたLLMを複数用意し、それらにメインのLLMが問いかけを発して合議制で答えるというような複雑な仕組みも併用されている。これが導入された理由はLLM一般にハルシネーションという幻覚を起こして誤った内容を正しいかのように答える場合があるのでそれを抑止するためである。所詮は前節で述べたような仕組みで「ありそうな答え」を生成しているだけなのでこのようなハルシネーションは避けがたい。

このような複雑な設定を行ったうえでChatGPTが公認会計士の短答試験をパスできるかを調べた。まず、今回は個々の問題の正否を問う設定にしてしまったのであてずっぽうでも5割は正解できる設定になってしまったこと気を付けよう。つまり5割より有意に高い正答率を達成しないとイケない。まず一番簡単な「正解は正解、間違

いは間違いと正しく答えた割合」を考えよう。ChatGPTには無料で精度が低いGPT-3.5と有料で精度が高いGPT-4がある。平成四年度の第一回、第二回それぞれについて調べたところ、GPT-3.5は2回中一回、GPT-4は2回とも60数%の正答率を達成したもののあてずっぽうに比べて有意に正答率が高いと言えたのはGPT-4を用いた、第一回目の試験だけであった（第1, 2回を合わせて一回の試験とみなすとGPT-4はあてずっぽうよりよいと言えたがGPT-3.5は言えなかった）。更に他の指標も調べた。例えば正しい選択肢のうち、正しく選ばれたものの割合（この場合、正しくない選択肢を選んだかどうかは一切カウントされない）という指標ではGPT-3.5による2回目の試験だけがあてずっぽうより有意に正答率が高くなった（第1, 2回を合わせて一回の試験とみなした場合は、単純な正答率とは逆に、GPT-3.5だけがあてずっぽうよりましだと判定された）。また、正解として選ばれた選択肢のうち、どれくらいが正解だったか（この場合、正解として選ばれなかった選択肢の正誤は評価から除かれる）の割合では、GPT-4の一回目の試験だけがあてずっぽうより有意にましだと判定された（第1, 2回をまとめて一回の試験と見た場合にはGPT-4だけがあてずっぽうより有意にましだと判定された）。このようにChatGPTによる短答試験の成績はばらつきが多かったので以下のような方法で仮想的に点を計算して合格点に達していたかどうかをみた。

個々の選択肢の正誤しか問うていないので、4つの選択肢のうち正しい2個を選んだら5点でそれ以外は0点という計算はできない。そこで、まず5点を1/4ずつ四つの選択肢に割り当て、正しい選択肢を選んだ時、および、正しくない選択肢を選ばなかったときにはこの1/4の点を加点することにし

た。この場合、ChatGPTが四つの選択肢のうち3つが正しいと判断すると、実際には0点のところ、もし、この3つのうち2つが正しい選択肢であれば、「正解を2つ正しく選ぶ、間違いを1つ正しく選ばなかったので4つの選択肢のうち3つは正解だった」とみなして5点の3/4の点を加える。なのでかなり甘めの採点であることに注意しよう。このように採点するとGPT-3.5は駄目だったが、GPT-4の方は第1, 2回とも65点を獲得した。これは実際の合格ラインである68点と73点に遜色がなく、合格相当と [3] では結論付けられている。つまり、ChatGPTは試験レベルでは公認会計士に匹敵する能力をすでに持っていると言えよう。

## 生成AIの計算力

生成AIであるChatGPTは現状でも公認会計士の試験をパスしてもおかしくないだけの性能を誇っていると解った。だが、ChatGPTは計算ができるだろうか？いくら試験にパスしても計算能力がなくては公認会計士は務まらないだろう（実際には電卓やPCソフトを用いるので計算ができないと会計士が務まらないとはいえないが）。

答えから言うと [4]、ChatGPTは桁が大きくなると単純な四則演算もおぼつかなくなる。これはChatGPTができることは「1000×4000を計算せよ」という文章を文字列として認識して学習した内容からもっともらしい答えを返すことだけだからである。桁が大きくなればそのような計算をした文書が少なくなってくるし、類推も難しくなる。極端な話、4桁の九九の表を覚えていないかぎり、ChatGPTは正しい4桁の四則演算はできないのだ。これはかなり致命的な問題なのでこれだけでもLLMが会計士の仕事をすべて奪うのはかなり困難だと言えらるだろう。

但し [3] では明示的に四則演算のルールを提示し、それを覚えさせ、それに従って計算するように指示すればできるようになるのではないかと示唆されている。これは一般にLLMがハルシネーションを起こして誤答を答えてしまうような場合でも、途中の考えを出力させると正解率があがることからあり得る話だ。途中を提示させるとその途中と矛盾した結論が結合した文章は人間が作った文章には起きにくいのもっともらしい解答を生成するという生成AIの本質上、誤答が減るから、と考えられている。

また、文字だけを扱うLLMあたかもが絵を描けるように振舞えたように、計算部分だけは計算に特化したAIに任せるようにすることで見た目上計算が正しくできるLLMを作ることも可能だろう。現状では計算ができないというだけですでに人間の公認会計士を置き換えることはほぼ不可能だろうが、この問題は比較的短期に解決される可能性があることを指摘しておく。

## おわりに

それでは結局のところ、生成AIは会計士の仕事を奪うだろうか？その答えはYESでもあり、NOでもある。YESという意味は、公認会計士が作業としてやっていること、例えば、いろいろなデータを集めて一個の表にまとめると言った作業はかなり複雑でも生成AIであるLLMに置き換えられてしまう可能性が高い。現状でも生成AIは画像としてしか存在しない表から中身を読みとって要約するみたいなことができるのである。紙の書類をかたっぱしからスキャンして生成AIに渡し「今月の収支一覧を作成しろ」と言ったらできてしまう、というのはそう遠くない未来に実現してしまう可能性が高い。

一方で、ハルシネーションなどがあるために、生成AIが作ったレポートのチェックは人間がやらなくてはならないだろう。その意味では答えはNOだ。新人やアルバイトに任せていた部分を生成AIに任せられるという意味では仕事は奪われるだろうが、(真の意味での)管理職的なポストはむしろ重要になるだろう。部下がいなくても部下がいる場合に匹敵するパフォーマンスを出せる可能性があるからだ。その意味では生成AIが会計士の仕事をすべて奪うことはないとはいえ、実際にはごく一部の生成AIを使いこなして効率をあげられる会計士だけが残り、有象無象は解雇されるという未来はありえなくもない。

その意味でも、拙稿をお読みの皆様には是非 [2] にアクセスして頂いて、実際に生成AIを試していただきたいと思っている。年配者こそ生成AIを使いこなして現役に返り咲くチャンスだと僕は考えている。みなさんの奮起に期待したい。

## さらなる読み物

---

より詳しい、しかし、専門家向けではない一般向けの読み物を紹介しておく。

- ・岡野原大輔 大規模言語モデルは新たな知能か 岩波書店 1540円 2023年  
ISBN 9784000297196
- ・スティーヴン・ウルフラム著、稲葉通将監訳、高橋聡訳、ChatGPTの頭の中、早川書房、1012円、2023年、ISBN 9784153400092
- ・清水亮、教養としての生成AI、幻冬舎、1034円、2023年  
ISBN 9784344986992
- ・小澤健祐、生成AI導入の教科書、ワン・パブリッシング、1760円、2023年  
ISBN 9784651203799

## 参考文献

---

- [1] Bing Chat <https://www.bing.com/search?q=Bing+AI&showconv=1>
  - [2] 学術講演会の記録 <http://www.granular.com/gakujutsu/>
  - [3] 増田 樹, 中川 慧, 星野 崇宏, ChatGPT は公認会計士試験を突破できるか? : 短答式試験監査論への挑戦, 人工知能学会第二種研究会資料, 2023, 2023 巻, FIN-031 号, p. 81-88  
[https://doi.org/10.11517/jsaisigtwo.2023.FIN-031\\_81](https://doi.org/10.11517/jsaisigtwo.2023.FIN-031_81)
  - [4] 湊 真一 ChatGPT はなぜ計算が苦手なのか  
[https://www.nii.ac.jp/event/upload/20230707-03\\_Minato.pdf](https://www.nii.ac.jp/event/upload/20230707-03_Minato.pdf)
-