テキスト生成 AI に関する技術論的考察

佐野正博、Sano Masahiro（明治大学名誉教授）

１．はじめに

ChatGPT など最近の生成 AI 技術の歴史的発展は、社会や生産のあり方を大きく変える可能性が高い。

ここでは ChatGPT などのテキスト生成 AI 技術に焦点を当て、技術論的視点からの構造分析に基づき、その可能性と問題点について考察をおこなう。

２．テキスト生成 AI の技術的構造

テキスト生成 AI の飛躍的性能向上をもたらしてコア技術は、Transformer 言語モデルである。(Vaswani 207; Uszkoreit 2017)

言語モデル（Language Model）とは、「文章や単語のパターンを学習し、自然な文章を生成したり、入力されたテキストに対して意味のある応答を返したりする人工知能の仕組み」のことである。

初期の言語モデルでは、「単語同士の結びつき」（単語同士の照応関係）を基本的対象としていた。

これに対して Transformer 言語モデルでは、各単語データに対して「文全体における単語の位置情報」を付与することで、「入力文章内の照応関係（類似度や重要度）」や「異なる文章同士の照応関係（類似度や重要度）」などを計算し、文脈の把握がなされている。

すなわち、「ある特定のコンテクストにおいて、ある特定の単語の次に、どのような単語がどの程度の出現確率で登場するのか？」という単語出現確率に関するデータベースが用いられている。

そのようにすることにより、「文章全体の結びつき」（文脈）をも処理対象とすることが可能となり、人間が書いたような自然なテキスト文を生成ができるようになった。

言語モデルに関する基本的性能指標の一つが、言語モデルの容量や複雑さを示すパラメーター数である。パラメーター数は、2018 年以後、1 年間に約 10 倍という飛躍的な増大を遂げている。OpenAI の GPT-4 言語モデルのパラメーター数は推定で 5,000 億~1 兆と言われており、その構築に巨額な費用が投じられている。

生成 AI は、そうした大規模言語モデルを利用して文脈を反映した単語間の連関確率を計算し、利用者が与えたプロンプト文に対応した回答を生成している。

文脈を反映した連関確率の数値などを含む言語モデルの形成に際しては、様々な先行著作物（public domain の茶作物やオープン利用が可能なネット上の各種データなど）を「学習データ」(training data)として利用するとともに、生成したテキストの適切性に対する評価をフィードバックする「教師あり学習」（Supervised Learning）などの機械学習によりテキスト生成の精度が高められている。

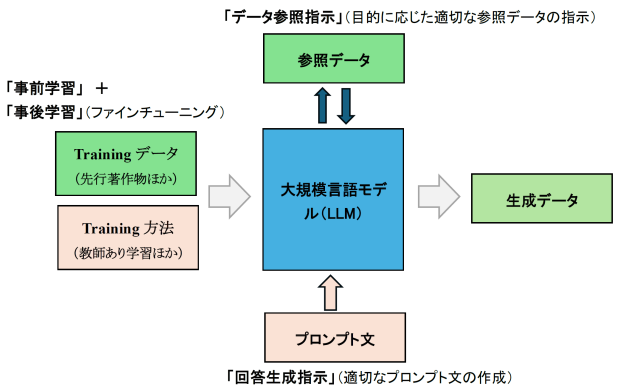


図 1. テキスト生成 AI の技術的構造

注) 筆者作成

２．生成データの質に関わる基本的規定要因

テキスト生成 AI の基本的な技術的構造は図 1 のようなものであるため、生成データの信頼度・妥当性を規定している主要な要因は下記の 4 つである。

1） 事前学習、および、事後学習・ファインチューニングにおける Training データの質と量
2） 事前学習、および、事後学習・ファインチューニングにおける Training 方法の質と量
3） 回答生成における参照データの質と量
4） プロンプト文の優劣

したがって生成 AI の回答の信頼度・妥当性を高め、社会的に有用なツールとするためには、Training データや参照データの質と量を高める必要がある。

そのための方策としては、著作権問題に配慮した上で、国立研究開発法人科学技術振興機構の J-STAGE 収録の学術ジャーナル、、日本の政府統計データ e-Stat(https://www.e-stat.go.jp/stat-search/)、国立国会図書館デジタルコレクション(https://dl.ndl.go.jp/)、国立公文書館デジタルアーカイブ(https://www.digital.archives.go.jp/)などの良質な資料やデータで生成 AI を学習させることや、生成 AI のデータ生成に際してそうした資料やデータすべてを簡単に参照できるような仕組みをつくることが重要である。（さらにまた生成 AI の回答の依拠データに関する traceability を確保して、回答の妥当性を簡単にチェックできる仕組みを組み込ませることも重要である。）

3. パターン認識に基づき「新結合」を実行するマシンとしての生成 AI - 生成 AI の「生成物」の Originality、Creativity、Innovativity 問題

先に論じたように、生成 AI は、「入力文章内の照応関係」や「異なる文章同士の照応関係」などに関する連関確率に基づき、テキストを「機械」的に自動生成するものである。

すなわち、人間の文章作成作業とは異なり、「内容的理解（意味了解）」や「理論的理解（概念的理解）」抜きに、「単語」を言語モデルに基づき単に「機械」的につなぎ合わせているだけである。

とはいえ、一昔前の自然言語処理マシンとは異なり、現在の生成 AI は、あたかも人間が作成したかのような極めて「自然」な文章を生成する。

こうしたことが可能となっているのは、大量のテキストデータを用いて「文章全体の結びつき」（文脈）を処理対象としたディープラーニングなどの学習をさせているからである。

そうした学習により AI は、テキスト、画像、音声などの大量データの中から、特定のルールやパターンに沿ったデータ構造を認識し、特徴抽出を行うなどパターン認識をおこなっている。

そして認識されたパターン（多様な現象の中に潜む共通構造）に基づいて、新たなテキスト、画像、音声などを生成している。

すなわち、生成 AI は、対象の中に潜むパターン（構造的連関）をディープラーニングなどによって捉え、そのパターンに基づいて既存要素を「新結合」させることによって新たな生成物を生成している。

シュンペーターの「新結合」論的イノベーション概念把握によれば、新発明がなくても「既存要素に関するこれまでにない新たな結合」によってイノベーションが生じる。

そのように既存要素の新結合によってイノベーションが生起するのであれば、生成 AI による既存要素の新結合によってイノベーションが生じても不思議ではない。

生成 AI の生成物の originality, creativity、innovativity の問題は、シュンペーターの「新結合」としてのイノベーション概念との関連で捉えなおすと興味深い。

注記　本稿に関連するより詳しい議論・参考文献は筆者作成の下記サイトを参照されたい。

佐野正博「生成 AI に関する技術論的研究」
https://ai.sanosemi.com/

生成 AI のインパクトと企業におけるその活用

――懸念・躊躇から「負けないため」の活用へ、そして残る課題――

那須野公人、NASUNO Kimito（作新学院大学名誉教授）

１．はじめに

2022 年 11 月 30 日のオープン AI による ChatGPT 発表以降、自然言語での指示にしたがってテキスト・画像等を生み出す、生成 AI ブームとなっている。それを象徴的に示すのは、生成 AI に不可欠な半導体 GPU のシェア 8 割を占めるエヌビディアの株価高騰である。また、日本経済新聞のサイトで“AI”という用語が使われている記事件数を検索してみると、2023 年からその件数が急増しており、AI 研究者の間でも、これを第 4 次 AI ブームと捉える見方が主流となっている。

ブームのきっかけをつくった ChatGPT は，大量データの学習による精度の向上、コンピュータの能力向上による素早い回答、自然言語対応による使いやすさ（「AI の民主化」）、幅広い分野の多様なタスクをこなせる汎用性で、急速に社会に広がっており、企業活動に大きなインパクトを及ぼす可能性が高い。

２．日本企業における生成 AI の導入

ChatGPT 発表後、その導入・活用を試みる企業が出始め、2023 年 3 月マイクロソフトがクラウドサービス上で提供を開始すると、入力データが AI の学習に使用されなくなったことで、それまで情報流出を警戒して導入を躊躇していた企業の背中を押すことになった。

AI を活用したサービスの開発を行っている㈱Exa Enterprise AI は、23 年 4 月・8 月・12 月に、自社の生成 AI 関連セミナー参加者を対象にアンケート調査を行った。4 月に最も多かったレベル 3 の「試しに使用」(43.0%)は、回を追うごとに減少し、他方レベル 5 の「日常的に使用」が、4 月の 7.2%から、8 月には 20.3%、12 月には 31.5%へと急速に増加して、これにレベル 4 の「時々使用」を加えると、71.2%にも達していた。

同社は、業務で積極的に利用する層が着実に増えており，多くの企業が試用から全社や部門での本格導入に移行したとみている。

PwC コンサルティングも、2023 年 10 月に生成 AI に関する実態調査を行っている。それによると、企業は「競合他社に先を越される可能性」や「新規競合の参入の可能性」を特に脅威と捉えていた。同社は，23 年の秋時点から見て、今後 1 年間が「本格導入」フェーズの 1 つの山を迎えるものと予見する。

また、生成 AI 導入による生産性向上に関して、23 年 2 月マイクロソフトのクラウドサービス上の GPT3.5 で、社内 AI "ConectAI"をいち早く導入したパナソニックコネクトは、ホワイトカラーを中心とした仕事は、①情報収集、②情報の整理、③ドラフトの作成、④仕上げ、という 4 ステップがあるが、①から③までは AI に任せることができるので、④の時間を増やすことができる。その結果、質の向上も期待できると述べている。

このように、日本企業における生成 AI の導入は、大手企業や IT 系を中心に、今や「本格導入」段階へと移行し、各企業は「他社に負けないため」に、特にホワイトカラーの生産性向上を狙い、生成 AI の導入を進めている。

３．生成 AI 発展の技術的背景

AI は、ディープラーニングモデル、データ 、ハードウェアが整うことによって、性能が格段に向上していった。とはいえ、画像認識の分野ではいち早く成果が出たが、自然言語処理の分野では精度が向上しなかった。その壁

を打ち破ったのが、2017 年にグーグルの研究者らによって発表された「トランスフォーマー」であった。これは、モデルサイズを大きくしていくと、ある段階で小型モデルにはなかった能力が現れ、できなかったことが突然できるようになることも分かってきた（Scaling Law〔スケーリング則〕）。その結果、生成 AI の開発は大規模化へと進んでいった。

生成 AI の中には、誤情報や非道徳的なコンテンツを生成するものもあり、批判の集中によって発表後数日で利用を中止するケースもみられた。ChatGPT が急速に普及したのは、時間とコストをかけて、非道徳的なテキストを生成しないようにファインチューニング（微調整）を行うとともに、人間に好ましい文章を生成するよう強化学習を繰り返したためであった。その結果、ビジネスで活用できるほどに、精度が高まってきたのである。

４．生成 AI の課題とリスク

生成 AI のリスクとしては、ハルシネーション（幻覚）、個人情報・機密情報の流出、著作権の侵害、意図的な悪用等があげられる。米アブノーマルセキュリティ（インターネットの安全対策を手がける）は、ChatGPT 公開後 2023 年上半期に、ビジネス詐欺メールが 89％増加したとしている。また、IBM の実験では、詐欺メール作成時間が生成 AI の活用によって、5 分に短縮したという（日本経済新聞電子、2023 年 12 月 31 日）。

しかし他方では、大規模言語モデルにおいても、リスク対応が進んでいる。

５． 生成 AI の雇用に及ぼす影響

代表的な研究のひとつ、ゴールドマン・サックスの“The Potentially Large Effects of Artificial Intelligence on Economic Growth.”（2023 年）は、世界的に 3 億人のフルタイム雇用に相当する人々が、自動化の対象にさらされる可能性があるとしている。しかし同論文は、自動化による労働者の移動は、歴史的に新技術にともなう新しい雇用の創造によって相殺されてきた。また生成 AI は、最終的に世界の年間 GDP を 7%増加させる可能性があるとも述べている。

さらに同論文は、アメリカにおいて自動化される雇用の割合として、事務・管理サポートを第 1 位（46%）、法律関係を第 2 位（44%）に挙げているが、肉体労働や屋外労働についてはほとんど影響がないとしていた。

今日では、このように比較的楽観的な見方が主流となっている。しかし、日本でも職務内容の変更によるリスキリングや、事務職等の大幅減少による雇用構造の変化への対応といった課題は残ることになる。

最後に、生成 AI とジョブ型雇用の関係について、若干触れておきたい。政府及び財界は、「失われた 30 年」からの脱却のための一手段として「ジョブ型雇用」に期待をかけている。しかし、これは 1910 年代のテイラー・システム、フォード・システムによる徹底的な労働の細分化に起源をもついわば古い制度である。だが、最近の欧米の実態調査によると、ジョブ型雇用のもとで作成される職務記述書は、抽象的・概括的な記述となっており、職務内容を柔軟化して事業環境への適応力を高める動きがみられるという。生成 AI の導入によってホワイトカラーの大幅な職務転換が予想されるなか、厳格な意味での「ジョブ型」の導入は混乱を生じかねない。2023 年から本格化した生成 AI ブームの影響は、「ジョブ型雇用」の議論にはまだ十分反映されておらず、生成 AI 時代を考慮したうえでの再検討が必要である。

資料と参考文献等は報告当日に提示します。

ワトソンＡＩによる賃金査定に関する問題報告

笹目芳太郎、SASAME Yoshitaro（JMITU 日本 IBM 支部書記長）

１．はじめに

2019 年 8 月、日本 IBM は AI（Compensation Advisor with Watson）を賃金査定に導入しました。会社は、組合が団体交渉で要求した AI の学習データや AI が所属長に示す情報の開示を拒否し、「AI が所属長に示す情報は、社員に開示することを前提としていない」、「あくまでマネージャーの判断をサポートするツールだ」との主張を繰り返しました。そのため組合は 2020 年 4 月 3 日に東京都労働委員会に救済申立を行いました。

AI 不当労働行為事件は、都労委の調査でも、また 2022 年 5 月の証人尋問でも、組合が要求したデータや情報は明らかにならず、翌 6 月からの和解協議でも和解に至りませんでした。そこで都労委の斡旋による和解協議を 23 年 12 月から継続し、ようやく 24 年 8 月 1 日に和解が成立しました。内容は組合側の勝利和解でした。

ここでは、AI（Watson）による賃金査定に関する問題について報告します。

２．AI を賃金査定に使うことの問題点

AI が労働現場で使われることには以下の４つの問題があり、とりわけ賃金査定に使う場合は、情報開示を行い、慎重に検討されるべきです。

①各人が自身で自分の個人データをコントロールする権利保障の問題

AI に考慮させたデータの中に、人種、信条などといった要配慮個人情報（個人情報保護法第２条第３項）などが含まれていないのかどうかを従業員に知らせなければなりません。これらの情報は賃金査定において必要のないものです。

GDPR（EU 一般データ保護規則）では、各人が自身で自身の個人データをコントロールする権利が保障されなければなりません。

会社は、賃金査定に使用している AI（Watson）は給与調整に当たって「各従業員について（2019 年度当時は）40 項目のデータを考慮し、4 つの要因（スキル、基本給の競争力、パフォーマンスとキャリアの可能性）ごとに評価した上で、具体的な昇給率を提案する」と説明しています。しかし、この 40 項目のデータは具体的に何であるかを明らかにしておらず、賃金査定において考慮すべきでない情報が含まれるのかどうかは不明です。つまり各従業員が自身で自分の個人データをコントロールする権利が保障されているとはいえない状態なのです。

②公平性・差別の問題

人間がもつバイアスによって AI の学習データや学習アルゴリズムが歪められることで、AI の判断（アウトプット）にバイアスが生じ、有害な結果につながる可能性があります。AI[の判断にバイアスが生じた場合には、個人および集団が不当に差別される恐れがあります。例えば、米国アマゾン社の AI を利用した人材採用システムは、女性を差別するという機械学習面の欠陥が判明し、2018 年に運用を取りやめました。また、機械学習においては、一般的に、多数派がより尊重され、少数派が反映されにくい傾向にあります（バンドワゴン効果＝いわゆる「勝ち馬に乗る」という現象）。 AI による賃金査定の場合にも同様の問題があります。

③透明性の問題（ブラックボックス化の問題）

AI はディープラーニングを行うためにアルゴリズムが複雑化しています。このため AI がどのようにアウトプットを生成したのかを誰も説明できない（ブラックボックス化）状況が生じます。AI による賃金査定の場合、従業員は AI アルゴリズムによって自身の個人情報がどのように 評価され、それによってどのような賃金査定の結果が生成されているのかわからない状況におかれます。査定の不透明性はＡＩ導入以前よりも高まります。

④自動化バイアスの問題

人間はコンピューターによる自動化された判断を過信する（自動化バイアス）という傾向があります。それ故、AI による賃金査定の場合、所属長は十分な検証をすることなく AI が生成した賃金査定の結果に従いやすくなる恐れがあります。

３．EU の AI 規制法案と労働者保護が進むヨーロッパ

EU（欧州連合）では、2021 年 4 月に公表した AI 規制法案で、人の安全や権利に影響を及ぼすリスクが高い AI を「ハイリスク AI」と分類し、適切な透明性確保を義務付けようとしています。これには、採用・昇進の決定、業務分配・業績評価等のために用いられる AI が含まれます。これは組合が要求している内容と同じです。AI 規制法は 2024 年に施行さ れ、2025 年までに適用が開始される見込みです。

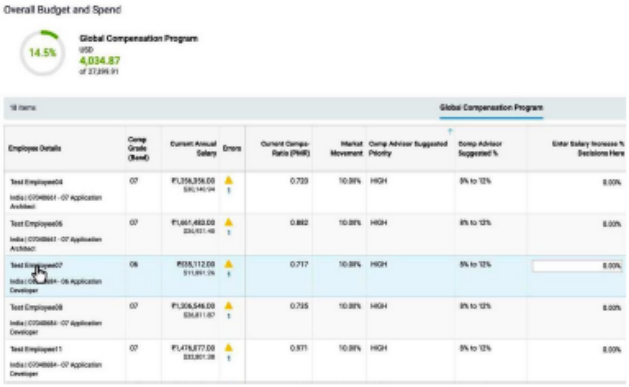
また、ヨーロッパの労働現場への AI 導入においては労働者の保護がかなり進んでいます。例えば、スペインでは使用者が自動化された意思決定システムを利用した場合、システム管理者は労働者とその組合に対して、以下を含む 相当な情報を提供しなければならないとしています。

・プログラム開発者とシステム導入者の名前
・システムの内容とその目的
・システムで用いられるトレーニングデータと変数の具体的情報

４．ワトソンＡＩによる賃金査定の実態

日本ＩＢMには賃金テーブルのような規定が無いため、各従業員の賃上げ額は直接の上司である所属長が決定します。

図は所属長が部下の賃上げ額を決定する際に入力する実際の画面です。左上にこの所属長が持っている賃上げ予算額（Overall Budget and Spend）が円グラフで表示され、その下に部下のリストが表示されます。部下はリスト上で、ワトソン AI が提案する賃 上げ率に応じたランク（Comp Advisor Suggested Priority）が高い順に、HIGH、MID、LOW の順に表示されます。所属長が各部下の昇給率（Enter Salary Increase % Decisions Here）を決定し入力すると、自動的に計算された昇給額の累計が左上の持ち予算額（Overall Budget and Spend）の円グラフにグリーンで表示され、持ち予算の残額が減っていく仕組みです。



この仕組みの問題点は、自動化バイアスにより、所属長はリストの上位の部下から順番に賃上げ額を決定していくためリストの下位の部下の昇給は無くなってしまう可能性があること、また十分な検証をすることなく AI が生成した賃金査定の結果に従いやすくなる恐れがあることです。

一方、部下には、ワトソン AI が自分の賃上げ率が何％と提案しているか、それによってワトソン AI が生成するリストで自分がどの位置に表示されるか、は知らされていません。

日本ＩＢMでは、給与調整は、所属長が格付規程５条に定める５要素（①職務内容、②執務態度、③業績、④スキル、⑤本給）に基づき行うことになっています。しかし、組合が 2020 年 4 月 3 日に東京都労働委員会に救済申立を行った際の証人尋問（2022 年 5 月 16 日）でも、会社側証人として証言した元人事労務担当ですら、ワトソン AI が考慮しているデータの全貌（令和元年度当時は 40 項目）を知らないこと、ワトソン AI が提示する内容と格付規程５条に定める５要素との関連性が不明確であることが明らかになりました。

こうして、ワトソン AI による賃金査定の運用実態の不当性は明らかです。

AI 不当労働行為事件関連文書やその他の資料、参考文献等については、報告当日に提示させていただきます。