

生成AIで進化する金融サービス／与える影響

藤井 達人

みずほフィナンシャルグループデジタル企画部の藤井でございます。本日はこのような場を与えていただきまして、大変ありがたく存じます。

一、生成AIとは何か

「生成AI」という言葉は、昨今新聞等々で度々取り上げられておりますので皆様も御承知かと思いますが、新しいものを生成するという点が、これまでのAIとの非常に大きな違いであると考えています（図表1）。新しいものを作るこ

とは従来のAIでもできましたが、非常にクリエイティブなものをAIが造ることに、とても大きなブレイクスルーがありました。さらに、テキストの作成が中心であったところ、ここ半年で画像や動画がかなりの精度で、かつ素早く作成できるようになってきていることが、非常にイノベティブです。

ChatGPTというサービスを御利用されている方も多いかと思えます。これはGPT(Generative Pre-trained Transformer)というテクノロジーを基にした会話型の生成AIサービスの一つで

図表 1

生成AIとは何か？

・生成AI（ジェネレーティブAI : Generative AI）とは？

- ▶ クリエイティブかつ現実的な全く新しいオリジナルのアウトプットを生み出す人工知能（AI）のことで、具体的には新しいデジタルの画像や動画、オーディオ（音声／音楽など）、文章やコードなどのテキストを生成するAI、もしくはこれらを組み合わせて生成するAIのことを指す。

・ChatGPT（Generative Pre-trained Transformer）とは？

- ▶ OpenAI社が2022年11月に公開した会話型AI。
生成AIの一種であるGPT-3.5/GPT-4を基に構築されており、教師あり学習と強化学習の両方の手法で転移学習されている。
- ▶ ChatGPTは、人間が明示的に○×をつけることで、好ましい対話が生成できるようになっている。

す。ChatGPTの裏には、人間が明示的に○×をつけるような作業を地道に行い、好ましい対話が生成できるようになっていることが特徴です。

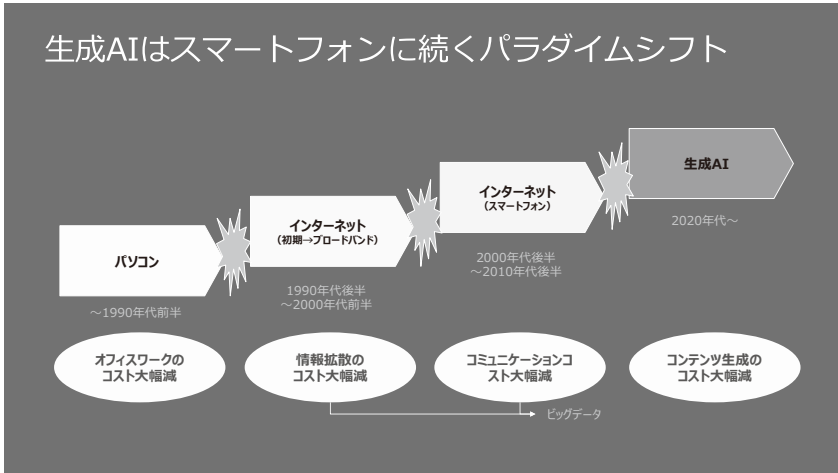
生成AIは比較的最近現れましたので、まだ実感が湧かない方もたくさんいらっしゃると思います。私どもは、過去二〇年〜三〇年のITの進化の歴史の中で、一つのパラダイムシフトが起きたと捉えています（図表2）。

振り返ると、まずパソコンが普及し始め、一九九〇年代前半頃には職場でも使われるようになりました。これにより、オフィスワークが非常に効率化しました。

次のパラダイムシフトはインターネットであると考えられます。最初はおもちゃのようにも思われていましたが、今となつてはビジネスになくてはならない存在です。

続いて同じインターネットですが、何といって

図表2



もスマートフォンです。モバイル端末で、人々が様々なことを完結し始めるようになりました。これもスマートフォンが発売された当初、特にWindows Mobileが出てきて、その後、iPhoneが発売されたあたりからすると、これほどまでに人々の中に普及してビジネスにも広く使われるようになるとは考えられなかったことですが、結果としてこれがパラダイムシフトになりました。

そして、次のパラダイムシフトは生成AIであると予想しています。人々の生活を大きく変えて、ビジネスも大きく変えていくだろうという仮説を持っています。当然ながら、二〇三〇年や二〇四〇年ぐらいにならないと、あの時の生成AIの登場がパラダイムシフトであったと認識されないかもかもしれませんが、少なくともそれぐらいのインパクトを持っているでしょう。そういう世界にどう対応していかないといけないのか、しつかり

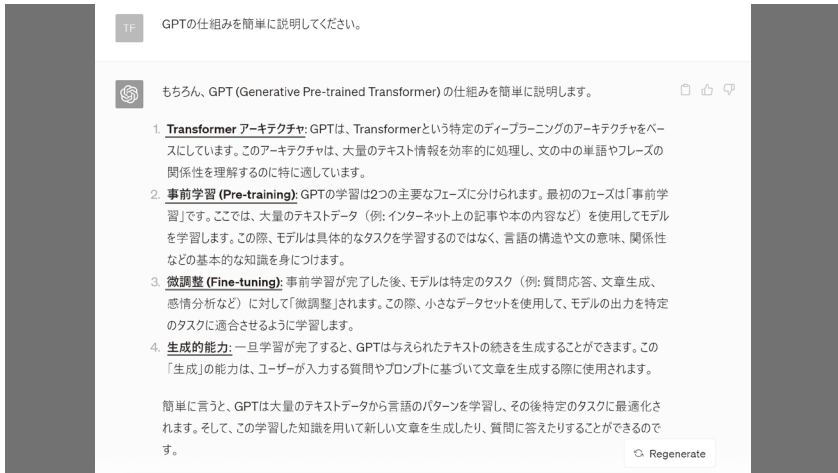
戦略に織り込んでいく必要性があるほどのものと捉えています。特に生成AIは、コンテンツをゼロから生成する一つ一つのコストが大幅に減少します。考え方にもよりますが、限界コストがゼロに近いということかなと捉えております。

代表的な生成AIのサービスとしてOpenAIという会社があり、同社は主に五つの生成AIの機能を提供しています。一つ目は、対話型のAIの開発です。GPT-3.5やGPT-4と呼ばれるもので、バージョンが上がるほど賢くなります。二つ目は、コード自動生成です。これは、プログラミングコードを自動生成するような機能です。三つ目はDALL-E（ダリ）という画像を生成する機能です。加えて、先ほどご紹介した会話型の生成AIのサービスであるChatGPT、音声を高精度のテキストに直すWhisperといったサービスがあります。これ以外にも動画を生成するサービス

や、音楽を生成するサービス、さらにはこれらをミックスしたものもあり、大変な勢いでこれらの技術が進歩していることが、今の業界の進化の速さを思わせるところです。

GPTの歴史を簡単に振り返ると、最新のバージョンはGPT-4と呼ばれるものです。これが登場したのは二〇二三年三月一四日で、既に八カ月程度経過しています。この世界では、八カ月というのはかなり長い期間になります。GPT-4が生成AIの中ではトップレベルの性能を持つモデル、エンジンであると言われていました。その前のバージョンのGPT-3もかなり賢いのですが、こちらは二〇二〇年ぐらいに登場しました。GPT-3、GPT-4がすごい星のごとく現れてChatGPTになったと思われるがちですが、研究開発自体はかなり前から行われており、段々と進化してきました。

図表 3



GPTはGenerative Pre-trained Transformerの略称です(図表3)。肝となるのはTransformerというアーキテクチャです。特定のディープラーニングのアーキテクチャをベースにしている、テキスト情報の処理や単語の関係性を理解することに適しています。もとはGoogleの研究者が発明したもので、この改良が飛躍的な性能の向上をもたらしたと言われています。Pは事前学習のことを指していて、Pre-trainingあるいはPre-trainedと言われていますが、GPTはPre-trainedと書かれているとおり、学習済みのモデルということです。地球上のネットに存在するありとあらゆるデータソースを学習に使っている非常に巨大なモデルがあります。そういったモデルが既に存在するという意味でPre-trainedと呼んでいます。

GPTのもう一つ特徴的なところとして、微調整(Fine-tuning)もよく聞かれる言葉です。こ

れは、先ほどの事前学習で学習済みとなったモデルが対応できない部分を追加で学習するようなイメージで、用途別にトレーニングできる機能があります。このような特徴を持つ GPT が登場して、AI に対する期待値が飛躍的に高まりました。

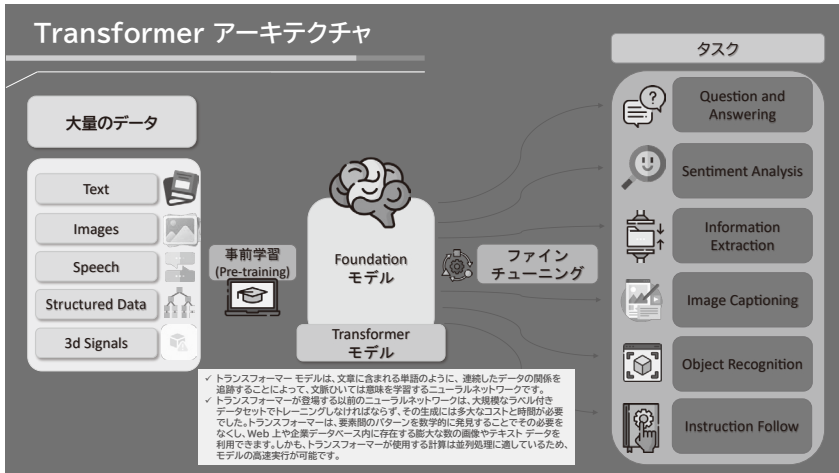
従来のモデルとの違いとしては、従来型の AI、マシンラーニングやディープラーニングをベースとする AI は、それぞれ目的とするタスクごとにモデルが作られていたと言われています。それに対して GPT がベースとする「Transformer のアーキテクチャは、ネット上に存在するあらゆるテキスト、モデルに合ったイメージやスピーチ等のデータを事前に学習して、基礎モデルあるいは基盤モデルと呼ばれる Foundation モデルを作ります。そこから微調整を施すことで、それぞれの目的ごとに適したモデルにすることが可能

になっています（図表4）。

GPT が得意とするタスクの例はこちらです（図表5）。これ以外にも非常に多様なことに対応できますが、いわゆる Pre-trained されたモデルを使ってどんなことができるのかと言われると、概ねこのようなタスクが特に業務上有効に働くと考えます。

例えば、要約やテキストを読み込ませて情報を抽出すること、あるいはスペルチェック、校正・編集などかなり得意としていてと思います。翻訳も、専用のサービスを使わなくても非常に高精度の翻訳ができます。分類系ではセンチメント分析、ちょっとした会話ログを読み込ませて、感情を数値で表しなさいという指示を与えると、比較的適切に感情を判定することも可能になっています。さらに、プログラミング系のコーディングも得意としています。弊社の中では VBA の

図表 4



マクロを開発するようなことを始めているケースがあります。開発の現場でも、生成 A I を組み込んだ開発ツールが今、本当に様々出てきています。将来的にはプログラマーはプログラムを書く機会があまり無くなるぐらいまで進化すると言われております。

1- ChatGPT の特徴

GPT というのは、どちらかというと単発で Q A をするようなサービスですが、ChatGPT は会話型で、いわゆる会話のコンテキストを引き継ぎながら更問いが可能になっており、まるで人と話し続けているようなサービスです。冒頭申し上げたとおり、ChatGPT は裏で人間がトレーニングデータに〇×をつけています。何故なら、ネット上のサービスをそのまま学習しただけでは何が誹

図表5

GPTが得意とするタスク例

大項目	タスク	説明
情報抽出系	キーワードベースの情報抽出	指定されたキーワードまたはフレーズに基づいて、テキストから関連情報を抽出
	事実の抽出	テキストから具体的な事実やデータを抽出
	要約	長いテキストから主要な情報を抽出し、短くまとめる
文脈理解系	文脈に基づく回答生成	ユーザーからの質問に対する回答を提供し、会話の文脈を理解する
	意図理解	ユーザーのテキストからその意図を解釈し、それに適切な応答を生成
	文法・スペルチェック	テキストの文法やスペルの誤りを検出し、訂正提案を行う
チェック系	校正・編集	テキストの構造やスタイルを改善する提案を行う
	内容チェック	提供された情報の正確性や矛盾をチェック
	言語間翻訳	一つの言語から別の言語へのテキスト翻訳を行う
翻訳系	異文化コミュニケーションの補助	異なる言語や文化のコミュニケーションを支援
	カジュアルな翻訳	スラングや俗語を含む非公式なテキストを理解し、適切な翻訳を提供
	センチメント分析	テキストの感情的なトーンを判断
分類系	テキスト分類	テキストを特定のカテゴリやトピックに分類
	トピック抽出	テキストから主要なトピックやテーマを抽出
	テキストタグ付け	テキストにタグをつけて、検索や整理を容易にする
文章生成系	コンテンツ作成	特定のトピックについての記事、ブログ、レポートなどを作成
	創作	ストーリーテリング、詩、小説など、創造的なテキストを生成
	カスタムレポート作成	特定の指示や要件に基づいてカスタムレポートを生成
	プロンプトに基づく作成	ユーザーからのプロンプトに基づいて、特定の内容を生成
プログラミング系	コーディング	プログラム自動生成、バグ修正、コメント生成

傍中傷で何が差別的なのか、AIは倫理的な判断ができないからです。そこで人間がきちんとした価値観を覚え込ませ、不愉快な思いをしないように注意を払っています。

このChatGPTに代表される生成AIは、質問力（プロンプトエンジニアリング）、即ちどのよう
にエンジンに指示を与えるかが結果の質を大きく左右します。ChatGPTの使われ方は御存知のように、同じ質問をしても同じ返答は無く、似たような質問をして少し詳しく指示を出すにより精度の高い返答が得られます。したがって、より良い答えが得られるような質問をするにはどうすべきか、工夫の仕方やノウハウ等も含めてプロンプトエンジニアリングと呼ばれています。聞き方が少し変わるだけでアウトプットがかなり変わってくるのは、少し人間的に思えるところもあり、科学というよりも芸術という表現が用いられるこ

ともあります。

従来のマシンラーニングに代表されるようなA Iは、通常だと同じ入力に対して同じ出力を生成しますが、GPTモデルの場合は確率論に従ってアウトプットを生成するので、ある意味では乱数のような働き方をします。ですので、同じプロンプト（入力する指示や質問）を与えても、どのような反応が得られるか予測することは困難であり常に一定の出力になるとは限りません。一部の最新のサービスでは、同じ入力に対しては一定の同じような答えを返す、少し固定するような機能も登場していますが、一般的にはChatGPTに代表される生成A Iは、同じ出力を返すことはありません。可能な限り詳しく、ステップ・バイ・ステップで指示を与えることで、できるだけ回答の質を平均化、平準化するということが、テクニクとして行われています。

プロンプトエンジニアリングは、幾つか類型化されたものがあります。例えばPrompt Process⁵⁾は、情報のアウトプットを見て生成A Iの理解が足りないなど思ったときに、それを補うような質問をする、あるいは質問の角度を変えるような形で、少しずつプロンプトを変えるものです。

Few-shot Learning⁶⁾は、例えば「ある質問に対してこのような回答を返してください」と、QとAのセットをあらかじめ例示して学ばせる手法です。幾つか例示することで回答のフォーマットの安定性を上げたり、回答の内容の精度を上げていくことができます。

Chain of Thought⁷⁾は、生成A Iに「あなたはこの問いに対してこういう段階で考えて答えてください」という指示を出して段階ごとに考えさせるようにすることで、論理を飛躍させず段階を踏んで考えるようにするものです。人間もいきなり

ゴールを目指す、間を飛ばし過ぎてうまくいかないことがあります。人間に優しく促すような形で、段階ごとに分かりやすく考えてくださいと指示を与えると、それに従って生成AIのモデルが考えて自分が欲しい答えに近いものが返るようになります。

これらのテクニックは、生成AIを使う側が、いかに自分が意図した答えが返ってくるかを引き出すテクニックです。先ほど申し上げたとおり生成AIのエンジンは同じ質問に同じ回答を行わないため、できるだけ回答を安定させる必要があります。そのための質問のテクニックがいわゆるプロンプトエンジニアリングです。今ご紹介したものの以外にも多くあり、日々新しい手法が考案されています。

一般的に、ChatGPTはユーザーエクスペリエンス、いわゆるコンピューターやネットワークに

対するユーザーの接し方を大きく変えていくのではないと言われており、私も同じ意見です。

ユーザーとしてはお客様や従業員が挙げられます。例えば、お客様がネットバンキングを使うときに、今であればウェブ画面を見て、ログインしたければログインのボタンを押す、残高照会をしたければ残高照会のボタンを押すといったように、ユーザーが画面を見てボタンを認識し、能動的にクリックして操作を行います。各社各様にウェブデザインは様々ですから、お客様がそれぞれのサイトの使い方を覚えて使い慣れることも必要です。従業員向けという点でも同じようなことが指摘できます。企業内に設けられているイントラネットの中から必要な情報を見つけるため、色々とクリックして、ようやく目的の情報にたどり着くということが度々あります。

しかし生成AIを使った時代のインターフェー

スになると、ひょっとしたらChatが基本となるかもしれません。「残高照会をして、もし余裕があれば定期預金を作成してください」のように、自然言語で指示をするイメージです。あるいは「従業員が引越をしたので住所変更したい」と入力すると、住所変更の方法を教えてください」ではなく「新しい住所を教えてください」と質問を受け、こちらが答えると自動的に住所がアップデートされることもあるかもしれません。このようなインターフェースが再定義されるきっかけになることも予想されます。

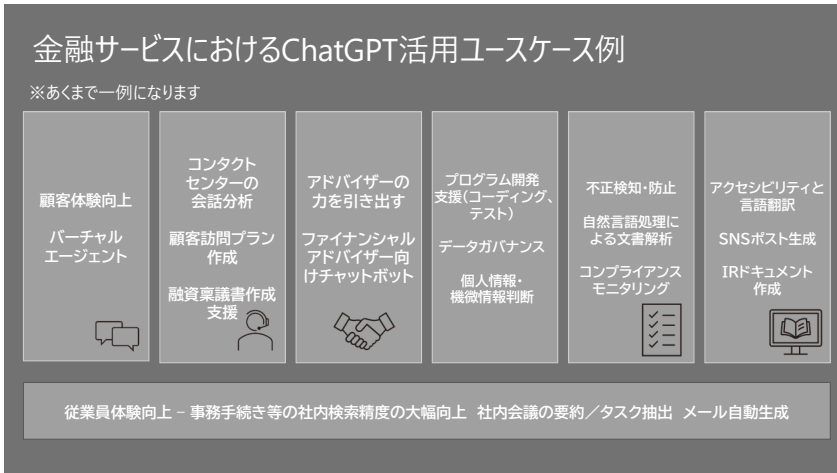
開発者もプログラミングコードを書く場面は減っていくと言われています。将来的に開発者は、プログラミングコードを書くよりは、プログラムの設計書をきっちり書くことで設計書から直接プログラミングコードが生成され、実行されるようになることも考えられます。もっと進むと、

プログラミングコードが何なのかは、あまり関係なくなるかもしれません。それぐらい自然言語とプログラムの距離感が近くなっていくでしょう。

三、生成AIが金融業界に与える影響

アクセントチュア社が発表した生成AIが業務に与える影響に関するレポートによれば、「生成AIは、現在の仕事を破壊し、ほとんどの労働者が『副操縦士』を持つことになる。人間とAIのコラボレーションの新しい次元を導入し、仕事の進め方や仕事の内容を根本的に変えるだろう」「ほぼすべての仕事に影響を受け、一部は廃止され、大半は変革され、多くの新しい仕事が生まれるだろう」と表現しています。また、業界別にAIによってどれぐらいの仕事が影響を受けるのかとい

図表6



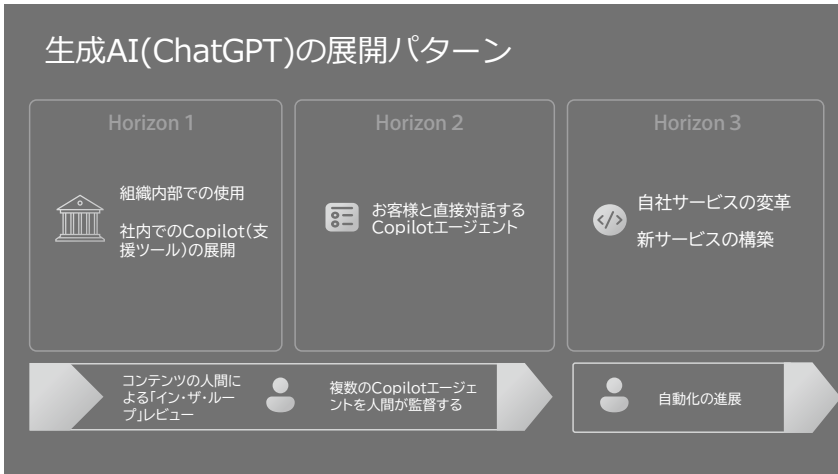
う予測も公表しています。それによれば、一位が Banking、二位が Insurance、四位が Capital markets となっていますから、金融業界がいかに生成 AI に影響を受けやすいと予測されているかがよく分かります。さらに銀行を例にとると、五四%の業務について Higher potential for automation、一一%について Higher potential for augmentation と評価されており、単純に過半以上の仕事が生成分成 AI によって大きな影響を受けるとされています。

金融サービスにおける代表的と思われる ChatGPT の活用ユースケース例はこのとおりです(図表6)。理想としては顧客体験向上やバーチャルエージェントのように、お客様に対して AI が直接対応することだと思いますが、まだ少し距離があるように思います。むしろ、コンタクトセンターでの会話分析や、それをもってコンタクトセ

ンターの担当者に、今回のアクションや顧客のニーズについてアドバイスをするなど間接的にお客様に対応することは非常にポテンシャルの高い領域であるように感じています。ファイナンシャルアドバイザー向けのチャットボットのようなサービスは、海外の銀行などで既に実用化しているとニュースでは出てきています。リスクマネジメントに関しては、例えば不正検知や不正防止、コンプライアンスモニタリングにも高い適用可能性があるということで、さまざまな研究が進んでいます。また、従業員の体験向上も一つのキーと考えています。弊社もそうですが、社内のイントラネットの使い勝手がものすごくいいかという点、必ずしもそうはなっていないという現状があります。こういったところを向上させるために、生成A Iは非常に大きく寄与すると思っています。

生成A Iの展開パターンについては、マイクロソフトがHorizon 1、Horizon 2、Horizon 3の三段階に分け定義しています(図表7)。Horizon 1である最初のステップは、組織内部での使用、あるいは社内でのCopilot(支援ツール)の展開です。副操縦士を英語でCopilotと言うように、完全な自動化には至っておらず人の作業の支援をするという位置づけの生成A Iの使い方 Copilotと表現します。したがって、全てを置きかえるというよりは、今の仕事のステップをできるだけ簡素にしたり、仕事の一部を自動化するといった使い方です。マイクロソフトを使用されている方は、Microsoft 365のCopilotというサービスをご存じかもしれません。そういった日常業務で使うようなツールの中に生成A Iが次々と取り込まれて、社員の日常の業務を支援していく、そのようなイメージだと思います。

図表7



Horizon 1で生成AIを使いこなす経験を得た後に、ステップ二であるHorizon 2として実際にお客様と直接対話するようなCopilotのエージェントをリリースすることになります。最終的にはHorizon 3であるステップ三として、自社サービスの変革や、新サービスの構築があります。例えば銀行では、様々な融資先、お客様企業との共創の関係の中で、新しいサービスを一緒に作っていくことが今後考えられます。

既にChatGPTを社内導入されている方々は、ChatGPTをそのまま使うと、企業の中のデータは参照しないため、企業内のデータを連携してサーチしたり、照会対象にしたいというニーズが非常に多くあると思います。この対応方法は様々ありますが、三点ご紹介します。

一つ目が、先ほど申し上げたプロンプトエンジニアリング、いわゆる聞き方をチューニングする

方法です。プロンプトに社内の情報を含めることで、その内容に対してQ&Aができるという非常に簡易的なやり方です。

二点目は、RAG (Retrieval-Augmented Generation) と呼ばれるものです。プロンプトエンジニアリングは非常に簡易的にできますが、参照できるドキュメントの量に制限がありますので、一般的に企業内データを照会する方法としてはこちらが主流です。これは、生成AIが、自分の持っている事前学習されたモデルの外にある、会社のさまざまな手続のドキュメントをデータベース化したものを参照しながら回答を作る方法です。社内のデータを使って検索したり、照会したいというニーズは、この手法で多くのケースに対処できると思います。

三点目は、先ほどご紹介した微調整 (Fine-tuning) で、生成AIのモデル自体にチューニング

を施すものです。モデル自体に会社の情報を読み込ませて知識を植えつける方法ですが、コスト面や一度学習した内容をクリアすることがかなり困難である等の課題もあるので、使うべき場面はまだまだ限られています。

四、生成AIがもたらす新たなリスク

ChatGPTには、大きな制約もまだまだあります。その一つが幻覚 (ハルシネーション) で、いわゆる堂々とうそをつくことです。特にHorizon 1、Horizon 2のお客様向けに出していく際には、生成AIを広範に展開していく上で克服しなければならぬ制限と考えています。

これは私がChatGPTに対して「日本にいた恐竜の名前を最低でも一〇個教えてください」と質

図表8

幻覚の例：

この例では、ChatGPTは存在しない架空の恐竜名をあたかも存在するかのように回答しています。

本来は確実に存在するものだけを言及すべきですが、ChatGPTは文章の意味や文脈を理解していないため、このようなことが起きます。

TF 日本にいた恐竜の名前を最低でも10つ教えてください。



日本で発見された恐竜の化石は他の地域に比べると少ないですが、以下は日本で知られている恐竜の名前の一部です。

1. フクイラプトル (Fukuiraptor)：福井県で発見された肉食恐竜。
2. フクイサウルス (Fukuisaurus)：福井県で発見された草食恐竜。
3. カモヤマティタン (Kamuysaurus)：北海道で発見された巨大恐竜。
4. ミカササウルス (Mikasaurus)：三笠市で発見された恐竜。
5. クボタルス (Kubotasaurus)：北海道で発見された恐竜。
6. シラミトリユス (Shiramitoriyus)：熊本県で発見された恐竜。
7. チュオサウルス (Chuoaptor)：福岡県で発見された恐竜。
8. タニハサウルス (Taniharsaurus)：宮崎県で発見された水生恐竜（実際には魚竜）。
9. タルボサウルス (Tarbosaurus)：石川県で発見された化石が所属する可能性が指摘されている恐竜。
10. タムラコケリア (Tamuracokeria)：岩手県で発見された肉食恐竜。

いくつかの名前は、正確には恐竜ではなく魚竜や他の古代の生物を指すことがあるため、恐竜に関する詳細な情報や分類を調べる際は、専門の文献や資料を参照することをおすすめします。

Regenerate

問した結果です（図表8）。ポイントは「最低でも一〇個」と聞くことで、そうすると律儀に一〇個答えます。しかし一〇個の中で正しいものは二つしかなく、あとの八つはChatGPTが適当につくり出したものです。本来は確実に存在するものだけを回答するべきですが、ChatGPTは文章の意味や文脈を理解して答えているわけではないので、こういうことが起きてしまいます。

最近では、答えてほしくないようなものや捏造したものを答えないようにするためのシステムの作り方のベストプラクティスが出てきました。それらを活用して、できるだけこのような現象を減らしていくことが一つの大きなポイントです。

また、先ほどの幻覚だけでなくその他にもリスクが挙げられます（図表9）。例えば「プライバシーと守秘義務」、「偽情報、毒性、サイバー脅威の拡大」です。来年にはアメリカの大統領選挙が

図表 9

生成AIがもたらす新たなリスク

リスク	説明
ミスと“幻覚”	生成AIが実在しないものを生成したり、人間同様の行動を取ることがある。これは、生成AIが言語の使用法を学ぶのみならず、世界の構造を十分に把握していないため、この問題に対処するためには、モデルの不確実性や限界について透明性を高める必要がある。
プライバシーと守秘義務	生成型AIモデルが学習データを記憶するという問題で、プライバシーや著作権の侵害へと繋がる恐れがある。具体的には、モデルが個人情報や機密情報を再現すること、著作権を有する素材を許可なく使用するなど。対策として、学習データの選定や保護に十分な注意を払うべきである。
偽情報、毒性、サイバー脅威の拡大	生成型AIが大量の偽コンテンツを生成・拡散し、デマやオンライン害悪の原因となる問題。人々の判別が困難となり、誤情報に騙される危険が増大する。解決策としてAI生成コンテンツの検出とラベル付けが必要である。
著作権への挑戦への時代	生成型AIが著作権素材を学習や模倣することで、著作権侵害の問題を引き起こす可能性があるという問題。モデルの行動がオリジナルの創作物を害する恐れがある。透明性や法的基盤の強化が解決策として挙げられる。
川下アプリケーションに響く埋め込まれたバイアス	生成型AIが学習データの偏見を増幅する問題である。この結果、不公平や差別的な出力が生じる可能性がある。学習データの選択と品質への注意が求められる。
価値観、整合性、良い指示の難しさ	生成型AIモデルが人間の価値や目標と一致しないかもしれないという問題について指摘。モデルが危険な結果を生む可能性がある。フィードバックと原則を用いてモデルの安全性を確保することが必要。

(出典) IMDA (シンガポール情報通信局) レポート “GENERATIVE AI : IMPLICATIONS FOR TRUST AND GOVERNANCE” より筆者翻訳・編集

行われますが、例えば生成型のAIが大量の偽のコンテンツを生成して拡散するというリスクがあります。具体的には、人の写真をアップロードすると、あたかもその人がしゃべっているかのような動画が作れてしまいますので、アメリカ大統領選挙の候補者が、言ってもいないようなことを言っているような動画を作り、それがSNSで拡散されてしまうことは危険性として十分に考えられます。そういったことが起きやすい時代になっています。それから、「著作権への挑戦の時代」です。知らないうちに生成AIが学習したデータの中に著作権に触れるものがあり、それが間接的に利用されることで著作権侵害であると思われるようなものが出てきてもあります。そのほかにも、「川下アプリケーションに響く埋め込まれたバイアス」は、例えば偏見を持つてしまようなデータを学習したモデルを使うと、その

モデルを使うアプリケーションが全て影響を受けてしまいます。生成AIがもたらすこういった新しいリスクも一方で認識しながら、正しいAIの使い方をしていかなければいけません。

これらが「責任あるAI」、英語で言うところ「Responsible AI」で、積極的な取り組みが必要になります。責任あるAIというのは、「安全で信頼できる倫理的な方法でAIシステムを開発、評価、デプロイするためのアプローチ」です。

企業が責任あるAIに取り組むべき理由は、生成AIの使用によって生じる潜在的なリスクを回避することが一番の目的です。先ほど申し上げたとおり、さまざまなAIが本来的に間違った意思決定を下す要素となるような情報を生成してしまう可能性が非常に高まっています。世の中にはアクセスしやすい生成AIはたくさんありますから、便利だからといってそういったものを気軽に

使って事業を行うと、どこかで取り返しの付かないことになる可能性もありますし、お客様に非常に御迷惑をおかけすることもあります。

生成AIを使うときは特に、①透明性を確保する。②AIによって出てきた情報に基づいて何かの意思決定を行う際は、それが説明可能であるようにしていく。③全てをAIに判断させるのではなくて、人間が最終的な判断を下す。④そもそも生成AIを用いる際に、倫理的な観点からそれを使っていいのか、その使い方は正しいのかをきちんと判断する、の四点が重要です。

加えて、そういうガイドラインや、そのガイドラインが企業の中できっちり運用されているかというモニタリングプロセスの定義や運用、さらには、各国で行われている議論をもとにして、政府や当局が設定するようなさまざまなガイドライン、ルール、レギュレーションの更新が常にあり

ますから、これらができるだけ迅速に取り込み、自分たちが責任あるAIの体制に回していくことが必要になります。したがって、生成AIをどんな使っていくという攻めの姿勢に対して、牽制するような動きをきちんと持たなければなりません。

少し観点が変わりますが、AIは環境にとっても悪いと言われることがあります。少し古いデータでは、AIのトレーニング、特に生成AIのトレーニングは非常に膨大なデータセンターのリソースを使い、自動車を動かすよりもはるかに大きなCO₂を排出されるとされています。特に、GPT-3、GPT-3.5、GPT-4といった最新のモデルをトレーニングしたり、推論エンジンを使うといった場合、使用するリソースは非常に高くなります。それはユーザー企業にとってはScope3と言われるCO₂の排出量が増加することになるの

で、AIを使えば使うほど環境に悪いこと、自分たちが使っているデータセンターが再生可能エネルギーで動いているかを意識してAIを使っていく必要があります。また、データセンターのハードウェアを冷やすための水も不足気味です。できるだけ低消費電力でAIを稼働するのに効率が良いコンピューター、半導体チップを開発していくことも大切です。ITプロバイダーに対しては、そういったさまざまな努力をもつてできるだけ環境を維持しながらAIを広めていく責務が生じているということも、知っておいて損はないと思います。

五、みずほフィナンシャルグループでの取り組み

弊社では、ChatGPTのWiz Chatという名称

で社内にも展開しております。今年六月から全社員に向けて公開し、制限なく全員が使いただけ使えるようにしています。今はGPT-3.5というバージョンで動いていますが、もうすぐGPT-4という最新のモデルが社内でも使えるようになります（注：その後、GPT-4が使用できるようになっていく）。今後さらに機能を拡張し、できることをさらに増やしたいと考えています。使い方をサポートする社内SNSのようなものを作り、そこで「こんなことができるか」という声に対して「こうやったらできる」というアドバイスをしたり、逆に社員から「こんな使い方でこんなことができた」という事例を共有してもらったり、社員がExcelを使ってプロンプトをつくれるツールを自発的に開発してSNSで公開したりと、ノウハウの共有が行われています。引き続き弊社では生成AIの開発を継続的に進めており、今年の六月

から七月にかけて生成AIのアイデアソンを実施したところ二〇〇件超の応募がありました。上位の案件は実装していく計画を立てているところ です。

後ほど申し上げますが、今の代表的な取り組みとしては、事務手続の照会対応のようなものを、Wiz Searchという名称で開発しています。それから、Wiz Createという、融資稟議の補助資料の基礎となる資料を自動的に作成するような仕組みの開発も行っています。まずフェーズ一として、とにかく生成AIを使ってみるためにChatGPTを導入しましたが、今は社内データや外部の最新データを活用するフェーズに移ってきています。

生成AI活用における体制の面では、先ほど申し上げたような業務に適用していく開発は業務適用推進チームが担っています。目先の業務への適

用のみならず、一〇年後に我々金融業界はどうなっているのかというところを見据えての戦略策定も行っています。また、生成AIを使ってどういうイノベーションが創出できるかを検討する創造的なワークや、経営陣向けの最新動向、業界への影響分析、啓蒙活動を担っているチームもあります。実際、生成AIは新しい技術や使い方が次々と現れており一カ月前のやり方よりもっと効率的な方法がある、という具合のスピード感で物事が進んでいます。それを支えるR&D (Research&Development) は、エンジニアを中心に最新の情報をアップデートしながら業務適用推進チームにフィードバックしています。一方で、AIの利用を野放しにしていると様々なリスクが顕在化するおそれがありますので、責任あるAIを推進するための、いわば守りのプロジェクトも始めています。まずはAI倫理基準のポリシーをつく

り、そのポリシーを基にしたガイドラインやモニタリングのプロセスを作成中です。

最後に、生成AIの開発の実際について、ご紹介します。

銀行には手続が山のようにあり、そのマニュアルも数多くあります。それらのマニュアルを照会するようなアプリケーションを作成しています。

ChatGPTはそのままですと社内の情報を読み込むことができませんので、先ほどご紹介したRAGという方法を使い、社内の事務手続のドキュメントをデータベースとして作りそれを生成AIと連携させて試しに動かす程度であれば、実は結構簡単にできるのです。

しかし、一見するとうまく動いているように見えるものの、現場の方に試しに使ってもらうと、粗が目立ちます。具体的には、回答の精度が低い、回答自体が間違っている、回答に参照した利

用文書のリンク先が全く違うなどです。こういった回答精度の低さを上げるために、いわゆるチューニングというか、システムの中で様々な工夫を行います。この時、利用者としてはできるだけシンプルに使いたいというニーズと、生成AI側はある程度指示をしっかりと与えた方がきちんと答えが返ってくるという裏腹の関係がありますので、バランスのとり方も意識して工夫しています。それから、GPT-4と言われる賢いAPIを使うと、遅延というほどではありませんが、少しレスポンスが遅くなることや、賢いAPIは単価が高いという課題もあります。

今後は、このような照会系のアプリケーションを利用者にとって最適化するべく、UIやロジックをチューニングする予定です。勿論、技術の進化が非常に速いですから、海外のソースや論文で使えそうなものがあれば試していきます。また、

Build or Buy という表現があるように、我々は開発に勤しんでいます。SaaS (Software as a Service) で使えるサービスがあれば、わざわざ作らなくても買ってきたらいいという選択肢も当然あると思っています。また、今後、社内のためにいくさまざまなデータは、いかにAIが読みやすいものになっているかを意識して、そういうフォーマットなりでためていく必要性も非常に重要だと思います。

まとめとしましては、ChatGPTに代表される生成AIは非常に大きな可能性がありつつも、まだまだ不完全な部分があるので、不完全であるということを前提にする必要があります。特に、責任あるAIの考え方のように、倫理的な問題や安全性の問題をきちんと担保して使っていくことが非常に重要です。Artificial General Intelligence (汎用型人工知能) がいつ実現してしまうかと今

盛んに言われておりますが、そういった世界観も見据えながら、我々として正しくA Iの技術を使っていくことにどんな挑戦していきたいと思っております。

私の話は以上とさせていただきます。御清聴誠にありがとうございます。

○森本理事長 藤井さん、最新の課題やテーマにつきまして、大変わかりやすく御説明いただいてありがとうございます。

それでは、皆様から御質問をお受けしたいと思っています。

○質問者A 企業内データを照会対象とするというところが、実務的に一番興味があります。三つ方法を紹介いただいて、プロンプトエンジニアリングやFine-tuningは割と単純なので分かりやすいのですが、RAGについてはある程度の仕組

みの作り込みが大変な気がします。ざっくりどのようなにお作りになっているのでしょうか。

○藤井 RAGに関しても様々なやり方があると思います。開発者の目線から、常に最新の環境を使って一番性能のいいものを使ってやりたいということになる、それなりの難しさがありますが、今、我々が行っているのは、クラウドのツールを使い、ある程度簡易な方法で対応しています。少し前は、ある程度知識がある人でなければできなかったのですが、GUIである程度できてしまうという最新のツールが出てきて、RAGを行う際のハードルは最近下がってきています。お勧めとしては、ある程度ツールの中で実現できるものを使ってやっていたくのがいいのではないかと思います。

○質問者A RAGの仕組みとしては、Large Language Model自体が検索するものではなく、

Graphical User Interface 経由でユーザーが検索していくことになるのでしょうか。

○藤井 具体的には、ユーザーが質問を入力すると、質問に対する回答を含む社内データはどこにあるかをまず探しに行きます。そして、そのデータを取ってきて、さらに Large Language Model を使って質問に答える、そういう二段階を挟んでいるとお考えいただけるといいと思います。

○森本理事長 生成 AI の利用は、企業や職場、日本全体で見て濃淡があると思います。御社では、職員に対する研修をどの程度幅広く行っているのでしょうか。また、まさに今、事務の改善が進行中なのですが、それを検討する、ある程度コアなグループやメンバーが形成されているのでしょうか。

○藤井 研修に関しては、使い方に関するレクチャーを紹介するビデオ会議を定期的に継続して

行っています。回を重ねるごとに少し工夫した使い方の紹介だったり、ユーザーがこういう使い方をしていると逆に紹介することで新しいアイデアをいただいたりと、インタラクティブな形も取り入れています。

もう一つは、先ほど申し上げたように、社内 SNS に専用のチャンネルを設けていて、そのメンバーになれば誰でも発言できて、誰でも聞きたいことが聞けて、誰でも答えることができるということがあります。我々が質問に答えなくても、社員同士でノウハウを持っている人がいれば勝手に答えてくれるような場をつくっています。さらに、今は準備中ですが、もう少し体系的な使い方をレクチャーするような動画コンテンツを社内の動画サイトに載せて、自由に見られるようにしたいと考えています。

担当メンバーに関しては、基本的に所管してい

るのは私どもデジタル企画部のメンバーですが、事務部門や、I T関係のグループ会社も加わっています。業務要件を出す人、業務要件を整理する人、実装する人等で一つのチームにしています。

今は段階的リリースをしていて、先月ぐらいに一番最初の初版を限定的に公開して、現場の人に評価してもらっているところです。評価というのは、具体的にこの質問とこの回答に対して五段階で何点か、あとコメントをつけるところもあるのですが、その回答に対して本当はこういう回答が欲しいというのを入れてもらい、そのデータをためて、それをもつてチューニングするというのを今繰り返しフェーズにいます。そんな形で、ウォーターフォールというよりは、どちらかというと少しアジャイル的に開発、改善ができるような体制を組んで、今やっております。

○質問者A 今、みずほフィナンシャルグループ

におられるので、今後はグループ内展開の仕方が出てくると思います。それについては何らかの考えはございますでしょうか。

○藤井 今は事務手続照会といった銀行のデータを対象にしています。グループワイドという意味では、何かやるようなユースケースが必要になりますが、これだということにはたどり着いていません。次に取り組みたいと思っているのは、様々なお客様にいただいたRFPに対する回答を生成するようなサービスについて、Fine-tuningを施してエンティティや部門ごとに適切なものが作れるようにすることです。

○質問者B G7あるいは外国の議会等でA Iの使い方についてガイドラインを作るという話がありますが、そういうものが有効なのかどうか、どのようにお考えでしょうか。

○藤井 生成A Iに対する確固たるガイドライン

やレギュレーションに相当するものはまだできておらず、まさに議論中です。ただ、G7でも、日本政府も、ガイドラインの制定に向けて様々な論点を出しています。私も金融業界でも責任あるAIに対してどう取り組むべきか、共通目線を合わせるための活動をする団体も出てきています。

世界的に見てもリスクがあるということははっきりしています。それに対して社会的にどういった取り組みやガイドライン・レギュレーションがあればうまく活用できるという共通認識にはまだ至っていないと思います。とはいえ、だからといって生成AIを活用し始めているのに何もなくていいのかというと、そうではないと思っています。様々な情報を収集しながら、我々なりにやっていくべきことを整理して、ガイドラインやモニタリングも検討しています。

○森本理事長 藤井さんには本当に最先端の話を

よく説明していただきました。ありがとうございます。
ました。改めてお礼を申し上げたいと思います。

(拍手)

(ふじい たつと・株式会社みずほフィナンシャルグループ)
執行理事 デジタル企画部 部長

(本稿は、令和五年十一月二十四日に開催した講演会での
要旨を整理したものであり、文責は当研究所にある。)

藤 井 達 人 氏

略 歴

株式会社みずほフィナンシャルグループ 執行理事 デジタル企画部 部長
1998年より IBM にてメガバンクの基幹系開発・運用、金融機関向けコンサル業務に従事。その後、MUFG のイノベーション事業に参画し、オープン API や MUFG コイン、アクセラレータプログラム等のプロジェクトをリード。KDDI の金融子会社にて執行役員 CDO としてモバイル金融に携わったあと、日本マイクロソフトの業務執行役員金融イノベーション本部長を経て現職。日本ブロックチェーン協会理事。東大 EMP 第17期修了。