米国における人工知能の進展

イーロン・マスク氏らが2015年に設立したOpenAI社が開発した「チャットGPT」。チャットGPTは基本的にはチャットボットであるが、その対応ぶりが人間そっくり、ということで話題となっている。さらに、チャットGPTは「生成AI」と呼ばれるように、ストーリー等を創作できる能力が注目されている。一方、米国では責任あるAIの運営についても重視されており、そのためのフレームワークも開発されている。

グローバルリサーチ研究所代表
青木　武
あおき・たけし　慶大卒。99年米ニューヨーク大学経営学修士（MBA）取得。信金中金総合研究所ニューヨーク駐在主任研究員等を経て、09年7月米国内にグローバルリサーチ研究所（IGR）を設立。http://www.igrlc.com/

1　チャットGPT

チャットGPTを利用してみると、従来型のチャットボットと比較するとかなりの進展を感じることができる。例えば、フォローアップの質問に答えたり、前提の誤りを見抜くことも可能であり、例えば「コロンブスは去年のいつごろ米国に来ましたか？」という意地悪な質問に対して、コロンブスが米国に来たのは1492年であって昨年ではないことを見抜くことができる。生成AIとも呼ばれるように、チャットGPTは過去のAIにはなかった創作能力も優れており、オーストラリアでは、チャットGPTが学校の課題のエッセイや作文などの宿題も簡単に作成してしまう、ということで禁止の動きがあるほどの人気である。名門ビジネススクールのペンシルベニア大学ウォートン校のオペレーションの授業でチャットGPTにテストを受けさせたところ、プロセスのボトルネックを探させたり、小売店の運転資金の必要額を質問したところ、的確に答えたという。結果として「B-」という、とりあえずは落第しないレベルの回答をしたという。具体的には、その一方で、コンピューターのくせに4桁の掛け算を正確に行うことができない場合がある。なお、現時点においてチャットGPTが学習しているデータは2021年までのものであり、例えば「明日の天気は？」といった質問には回答できない。

チャットGPTはAPIでも提供される。これにより、企業等は自らのソフトウエアにチャットGPTを組み込み、機能を拡張できるようになる。一方、この優れたツールを悪用することも可能であり、犯罪者グループ等がより洗練されたマルウエアを製造したり、フィッシング攻撃を行ったりすることも可能となる。

チャットGPTは金融分野でも利用が考えられる。ただし、課題を先に言及しておくと、AI共通の課題であるが、回答自体はおおむね優れているが、なぜそのような回答になったのかを説明できない。よって、監督当局向けに説明が必要なことに

は使いにくい。次に、チャットGPTに個人情報やセンシティブな情報を提供した場合に、当該情報が将来的にどのような利用をされるかが分からない、という問題がある。また、チャットGPTはよく間違える。後述する通り、大手金融機関はチャットGPTの利用を禁止している。そうしたことを踏まえた上で、筆者がチャットGPT自身に金融分野におけるチャットGPTの利用法を尋ねたところ、以下の回答であった。

マイクロソフトと「チャットGPT」のロゴ

・カスタマーサービス：AIチャットボットを作成してシンプルな課題に対応する。

・パーソナライズしたマーケティング：顧客データを分析してパーソナライズした商品のお薦めを行う。

・コンテンツの作成：電子メール作成やウェブサイトのコピーなどをチャットGPTに作成させる。

・不正防止：ID盗用やクレジットカード不正を見抜く。

・リスク管理：投資や融資における情報収集に寄与することにより、リスクを管理する。

このほか、米ニュースサイト「インサイダー」では以下のような金融分野でのチャットGPTの可能性について述べている。

1）投資銀行業務では、すでにスウェーデンのプライベートエクイティファンドであるEQTがチャットGPTを利用している。EQTでは、以前からマザーブレインと称するデータプラットフォームを構築しており、投資すべきベンチャー企業に関するデータを利用していた。一方、膨大なデータを検索する手間が課題であった。そこで、EQTではマザーブレインにチャットGPTの検索機能を組み合わせ、会話形式で簡単に必要なデータを得ることができるようにした。

2）投資については、ウォーレン・バフェット氏の投資戦略と、もし彼ならどのような株式に投資するか、と質問したことをインサイダーは紹介している。チャットGPTの回答は、同氏の投資戦略は市場から過小評価されているバリュー株のうち、今後の成長の可能性が高い株式に投資すること、と指摘し、その上でペプシコとユニリーバを同氏が投資しそうな銘柄として挙げている。

3）サイバーセキュリティや不正防止の分野では、例えば口座開設を申し込んできた顧客が正当な顧客なのかどうか、他のデータと照合して確認することが考えられる（ただし、先述の通り、個人情報を入力した場合の将来の漏洩リスクについて考慮する必要がある）。

4）リテールバンキングでは、バンク・オブ・アメリカのチャットボットであるエリカのように「先月アマゾンでいくら使いましたか？」といった質問に回答することも可能と思われる。また、顧客のエクスペリエンスをパーソナライズしたり、当該顧客に合った商品等をお薦めしたりすることも可能と思われる。もちろん、顧客が商品やサービスについて質問がある場合に回答することもできる。

一方、チャットGPTは回答が不確かな場合でも、人間がよくやってしまうように「わかりません」と言うのではなく、ウソと真実を巧妙に組み合わせることにより、うまくでっち上げてしまう、という問題も指摘されている。例えば、南部の大手銀行であるTruist銀行のAIヘッドであるBjorn Austraat氏は、チャットGPTに「鶏の卵は牛の卵より大きいですか？」と質問したら、「いいえ、牛の卵は20㌢ほどの長さがあり、5㌢ほどの長さの鶏の卵の方が小さいです」と回答したことをソーシャルメディアのリンクトインで紹介している。そこで、筆者も「牛の

卵の人気レシピを教えてください」と尋ねると、4種類の「牛の卵」を使った料理をでっち上げてくれた。しかしながら、その後無関係の質問をいくつかしてから、牛の卵に関する質問をしたところ、「すみません。さきほどは間違えました。牛は卵を産みません」と自分の間違いを訂正する学習能力を持っていることは驚きである。

なお、マイクロソフト社は、チャットGPTを同社の検索エンジンであるBingに組み込んでおり、検索のバラエティの幅が広がることになる。これは検索エンジントップのグーグル社にとっては大きな脅威であり、グーグル社も検索能力のあるチャットボット機能をリリースしており、チャットGPTと正面から対抗することになる。ただし、チャットGPTと同様にグーグルのAIも間違えると指摘されている。これは、AIが本当の意味で文章を理解しているからではなく、あくまで確率に基づいて回答しているためである。チャットGPTを運営するOpenAIは積極的に事業の拡大を図っており、グーグル等から人材を採用している。インサイダーによると、OpenAIはグーグル社から59人、メタ社から34人、JPMチェースからも7人の人材を雇用している。

一方、JPMチェース等の米国4大メガバンクおよびゴールドマン・サックス、ドイツ銀行は行内でのチャットGPTの利用を禁止している。禁止の理由は明確にはなっていないが、メガバンク等では、トレーダー等がコミュニケーションのためにWhatsAppを利用したことで合計で20億ﾄﾞﾙものペナルティを監督当局から課されたこともあり、情報漏洩リスクのあるサードパーティーアプリを利用することに慎重になっている。そもそも金融機関は新しいシステムを導入するかどうかを決める場合にサイバーセキュリティ等のリスク査定をすることになっており、導入するにはリスクが高い、と判断したものと思われる。

実際、チャットGPTについては、金融機関にとって幾つかのリスクがある。

①知的財産の侵害：OpenAI、マイクロソフト、GitHubはソフトウエアの著作権の侵害だとして訴えられていると伝えられている。

②セキュリティ：アマゾン社の弁護士は従業員に対し、機密情報をチャットGPTに入力しないように警告している。チャットGPTを運営するOpenAI社が機密情報をどのように取り扱っているかが不透明、ということである。

③②にも関連するが、特に金融機関は先述の通りソフトウエアやコミュニケーションツールの利用に関しても規制されており、多くの場合は履歴を保管して監査が受けられる状況にしておく必要がある。

こうしたことから、ネオバンクのチャイムはチャイム独自のチャットGPTのような生成AIの仕組みを構築中であるとインサイダーは伝えている。当該プロジェクトは、チャイムのプライベートクラウド内で知的財産を侵害しない形で学習を行う予定である。もっとも、当該プロジェクトは1月に始まったばかりである。それでも、今後数カ月で仕上げ、年内には実用化を目指している。当面はエンジニアにとって、開発が容易となるような生成AIを目指しているようだ。メタ社も生成AIのためのチームを立ち上げたとザッカバーグCEOが発表している。もともとはOpenAIの創業者の一人であるイーロン・マスク氏もチャットGPTのような仕組みを構築する方向で人材集めをしていると報じられている。チャットGPTによって開かれたパンドラの箱は、もはや元には戻らず、チャットGPTのような生成AIが成長分野であることは間違いなさそうである。

2　AI倫理の進展

このように機能面で大きく進展しているAIであるが、その倫理面も注目されている。22年10月にホワイトハウスは、AI権利章典の青写真と称するAI倫理に関するガイドラインをリリースした。これは法的効力があるわけではなく、銀行として順守の義務があるわけではないが、米消費者金融保護局（CFPB）をはじめ連邦金融監督当局がAI権利章典に基づき監督・検査を行う可能性は高いと思われる。特に、採用と融資審査におけるAIを名指しで念頭に置いているため、金融機関にとっては少なくない影響があると思われる。AI権利章典には五つの原則がある。

①国民を不安全で非効果的な自動システムから保護する：AIシステ

ムはテストが行われ、当初の意図通りに機能していることを確認するべきである。不適切または無関係のデータが利用されるべきではない。独立した評価と報告が行われるべき。

②アルゴリズムによる差別を防止する：人種、肌の色、性別、宗教、年齢、障がいの有無等で差別されるべきではない。アルゴリズム開発者はこうした差別が発生しないような継続的な対策を講じるべきである。独立した評価、テストの結果および問題緩和策についてわかりやすく開示すべきである。

③データプライバシーの悪用を防ぐ：ＡＩシステムにはプライバシー保護がデザインとして組み込まれている必要がある。個人のデータを収集・利用・移動・削除する場合は本人の承諾が必要である。現在行われているような広範で難解な合意文書ではなく、用途を特定して理解しやすいものであるべきである。特に、金融、健康、就労、教育、司法については、センシティブな分野であり、必要な機能に限定してデータが利用されるべきである。また、教育、就労、住宅について、継続的な監視が人権を侵害することがあってはならない。

④自動化システムが使われていることを利用者に周知し、説明する：自動化システムが利用されていることを利用者が知るべきであり、それがどのような影響を与えるかを理解すべきである。そのためのわかりやすい情報を開発者等は提供すべきである。自動化システムは技術的に確かな方法で、意味があり使いやすく、内容のリスクに応じてシステムを理解できるような説明を提供すべきである。

⑤自動化システムを利用しないことを選択できるようにする：自動化による決定を好まない場合、人間による決定のオプションも提供されるべきである。システムの決定に不服がある場合は、人間の担当者による判断にエスカレーションできる仕組みを提供すべきである。当該担当者は必要なトレーニングを受けた者であり、利用者からアクセスしやすい必要がある。特に司法、雇用、教育、医療の分野では必要である。

このうち、①および②は特に目新しいものではないが、②についてはテストの結果や差別対策等の開示まで求められている。大手銀行のウェルズ・ファーゴでは、ＡＩを開発したチームとは別のチームが、ＡＩのアルゴリズムやデータが適切であるかどうかをチェックしているという。③個人情報の保護については金融は具体的にセンシティブな分野として挙げられているため、より注意が必要と思われる。特に、個人情報を収集・利用する場合、利用目的ごとに承諾を得ることがより厳密に求められる可能性があるように思われる。ただし、融資審査等にＡＩが利用される場合は開示されている場合が多いが、マーケティングの場合は必ずしも開示がされていないことが課題であると思われる。また、チャットが人間によるチャットなのか、ロボチャットなのか明確ではない場合も多く、これは④に抵触する可能性があるように思われる。金融機関にとって悩ましいのが⑤であり、もし融資審査でロボが却下した場合、人間による再レビューを求められてそれに応じられるか、というと現時点では対応していない場合も少なくないように思われる。ウェルズ・ファーゴでは、ＡＩのスコアリングは利用しているが、最終的に申請者に融資をするかどうかは人間が判断するようにしているため、この問題は発生しないと考えているという。

さらに、米国立標準技術研究所（ＮＩＳＴ）は23年１月に国家ＡＩイニシアティブ法に基づき、ＡＩ管理のフレームワークを発表した。ＮＩＳＴは規制当局ではないため、これは強制や規制ではないが、金融機関や企業にとって今後のＡＩ管理の指針になると見られる。先述のＡＩ権利章典の青写真がＡＩ管理の一般原則を大まかに定めているのに対し、ＮＩＳＴのフレームワークは実際に企業等が行うべき作業等を定めている。ＡＩは急速に発展しており、その適切な運営・管理が不可欠となっている。ＡＩが普通のプログラムと異なるところは、ＡＩは学習し続けることにより、プログラムを自分で書き換えることができる点にある。よって、ＡＩを利用する金融機関や企業等はＡＩが正しい方向に書き換えられているのかどうかを監視していくことが重要となる。

ＮＩＳＴのＡＩ管理フレームワークは①統制②マップ（特定）③測定④管理―により構成されている。①統制はすべての作業の基礎になるものであり、全体的な方針、ＡＩを特

定、測定、管理する方法や手順等について規定や手順書で定める。また、組織内での責任の所在を明確にする。②マップ（特定）はＡＩについて関連する法令、必要なスキル、目的、業務上の利用方法、リスク許容度、利用するモデル、当該ＡＩ利用のメリットとデメリット、利用する範囲、予想されるインパクト、担当者・責任者等を特定する。③測定は、実際に測定する対象や測定方法を定め、ＡＩが当初の目的通りに機能しているかどうかを測定することである。ＡＩシステムは定期的に評価される。また、トラッキング状況は記録される。④管理は、ＡＩを特定したり測定したりするために必要なリソースを配分したり、実際に必要な対応を行うことである。例えば、不具合が発生した場合にそれを解決する。

ＡＩを利用している金融機関や企業はすでに何らかのＡＩ関連規定・手順書等を有しているはずであるが、それらを改定したりアップデートする場合にも同フレームワークは参考になると思われる。また、これからＡＩを利用することを検討している金融機関にとっては、ＡＩポリシーの策定に役立つと思われる。