3 サービスイノベーション総合研究所

“ひと”がWell-beingになるために、“ひと”とAIが協働する未来を創造

近年登場した大規模言語モデル（LLM：Large Language Model）は汎用的なAI技術として飛躍的な進化をとげ、様々な活用が検討され始めている。サービスイノベーション総合研究所は“ひと”とAIが協働する未来を創造するためのAI技術の開発を進めており、LLM技術もその一つの柱として研究開発を進めている。

日本電信電話株式会社
研究開発担当役員
サービスイノベーション総合研究所
所長　**大野 友義** 氏

すべての“ひと”がWell-beingになれる未来をめざして

IOWN（Innovative Optical and Wireless Network）におけるサービスイノベーション総合研究所のミッションとして、すべての“ひと”がWell beingになれる未来を創り出すことをめざしている。我々はこれまでに分散型で持続可能な社会において、NTT独自のサイバー・フィジカル融合技術によって実現すべきひとをWell-beingにする8つの世界観を打ち立て、デジタルツインコンピューティングによる社会課題解決と新たな価値観の創出を推進してきた。

昨年はNTT独自の大規模言語モデル（LLM）「tsuzumi」発表をはじめ、“ひと”とAIの協働社会を具体化するための技術の開発を推進した。NTT研究所が長年にわたって積み重ねてきた自然言語処理技術とAI技術が融合することで、軽量で高い日本語処理能力を有するLLM「tsuzumi」を実現した。

また、近年の巧妙なフィッシングサイトを、LLMを用いて検出して攻撃を防ぐ技術を開発した。このように、AIを活用したセキュリティ対策の技術開発も行っている。

そして複数のAIが連携することで高度な問題に対して多様な視点からの解決をもたらす「AIコンステレーション」技術の開発に着手している。登場以来進化し続けるAI技術の問題を解決し、より有用な技術とするための研究を行っている。

本稿では、“ひと”とAIの協働社会を実現するためのLLM技術に

図1　ひとをWell-beingにする8つの世界観と技術領域

ついて、紹介する。

NTT 独自の大規模言語モデル「tsuzumi」

NTT 人間情報研究所は、大規模言語モデル（LLM）の普及に伴い課題となっている電力やコスト増加などの課題解決に向け、軽量でありながら世界トップレベルの日本語処理性能を持つ大規模言語モデル「tsuzumi」を、2023 年 11 月に発表した。

・tsuzumi の特長

「tsuzumi」は、研究所が保有する 40 年以上に及ぶ自然言語処理研究の蓄積、世界トップレベルの AI 分野の研究力を活かし、その特長である軽量かつ世界トップレベルの日本語処理性能、柔軟性、そしてマルチモーダル対応を実現している。以下、これらの特長を簡単に紹介する。

・軽量な「tsuzumi」

「tsuzumi」は、パラメータサイズが 6 億の超軽量版と、70 億の軽量版の二種類ある。超軽量版は CPU で、軽量版は 1GPU で高速な推論動作が可能。これは、GPU クラウドの利用料金に換算すると、学習コストを約 300 分の 1（超軽量版）および 25 分の 1（軽量版）、推論コストを約 70 分の 1（超軽量版）および 20 分の 1（軽量版）に低減できる計算である。（NTT 試算）

・日本語が得意な「tsuzumi」

「tsuzumi」は日本語と英語に対応しており、特に日本語処理性能については長年の研究で得た知見を活かすことで、高い性能を実現した。 生成 AI 向けのベンチマークである Rakuda では GPT-3.5 や国産トップの LLM 群を上回ることを確認した。（図 2）

英語でも、世界トップクラスと同程度の性能を実現しており、多言語にも今後対応する予定である。

・柔軟な「tsuzumi」

LLM に新しい知識を追加で学習させようとする場合、膨大な数のパラメタ全てを再学習させると、計算にかかる学習コストが大きくなる。

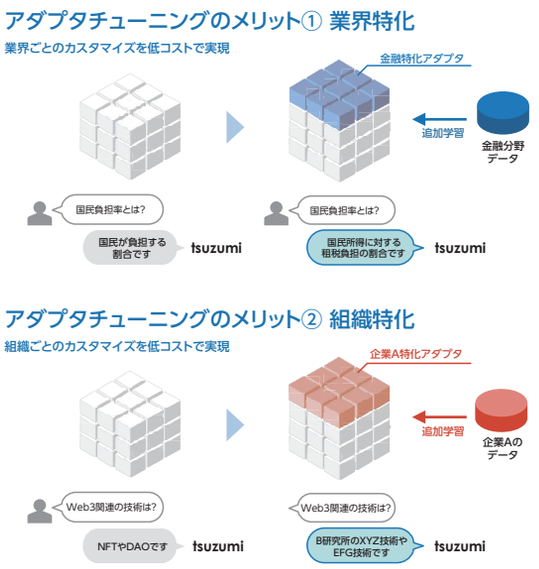

図 3　アダプタチューニング概要

そこで「tsuzumi」では、効率的に知識を学習させることができるアダプタにより、例えば特定の業界に特有の言語表現や知識に対応したり、組織ごとの知識に対応するようなチューニングを、少ない追加学習量で実現できる。（図 3）

・マルチモーダルな「tsuzumi」

言語化されていないグラフィカルな表示や音声のニュアンス、顔の表情などを理解し、現実世界での人との協調作業を可能とするような、マルチモーダルへの対応も予定している。

・今後の展開

現在は、人が AI に合わせていく使い方をしているが、人間情報研究所では、人と同等の入出力インタフェースを持ち、任意のマルチモーダル･身体行動タスクを実行できる、いわば人間社会に自然に溶け込む AI の実現を目指している。

この目標達成には、マルチモーダルの更なる深化や、AI の自律性や規範性といった AI ガバナンス面など、多くの課題が存在する。NTT 人間情報研究所は、これらの課題を

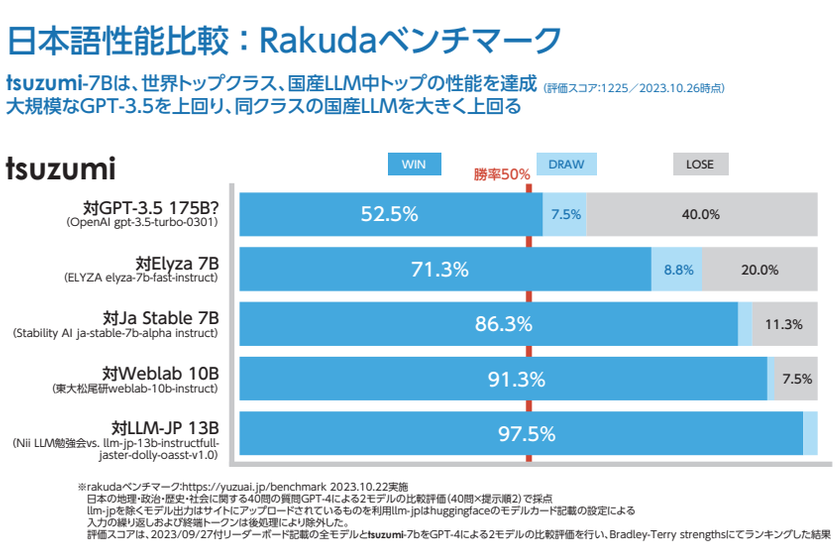

図 2　市中 LLM との日本語性能比較

解決するための研究開発を推進し、AIと協調した豊かな社会を実現することにチャレンジしていく。

LLMを用いたフィッシングサイト検出技術ChatPhishDetector

巧妙な手口でユーザにアクセスさせ、個人情報や口座情報などを盗むフィッシングサイトの被害は、件数・被害額いずれも増加傾向である。この状況に加え、近年ではサイバー攻撃者向けのAIサービスが登場するなど、フィッシング被害がさらに増大する可能性もある。

NTT社会情報研究所は、さまざまなソーシャルエンジニアリング攻撃の中でもフィッシングサイトのように人を騙すWebサイトの特徴把握や対策手法の創出に向けて従来より研究開発に取り組んでいる。最近のLLMの目覚ましい発展を受け、世界に先駆けてNTTセキュリティホールディングスと共同でLLMを用いた、“ChatPhishDetector”を開発した。

この技術は、これまでの研究成果に基づくフィッシングサイト判断ポイントを、LLMへの指示として与え、LLMの持つ高い言語処理能力で判断させるものである（図4）。

LLMに与える情報は2つのステップで作られる。1つ目のステップでは、フィッシングサイトの疑いがあるWebページを巡回し、「HTMLデータ」や「OCRでテキスト化した画像データ」を取得する。2つ目のステップでは、前ステップで取得したデータをLLMへの指示（プロンプト）に変換する。変換の際、1つ目のステップでテキスト化したデータは、LLMの入力文字数制限等の条件に合わせつつ、フィッシングサイトの特徴を残すように簡略化する。

LLMへの入力には、精度よくフィッシングサイトの判定ができるよう、段階的な判断手続きで処理するよう指示を与えたり、回答が文脈に沿うようなロールを与えたりする、プロンプトエンジニアリングの手法を用いている。

本技術について、LLMとしてGPT-4を活用した実験では、正規のサイト1000件、フィッシングサイト1000件について、98.4%の精度で判定できた。

ソーシャルエンジニアリング攻撃において人を騙す基本的なパターンは普遍的ではあるものの、具体的なユースケースやシナリオは多様化しているため、従来は機械的に判定するための個々の特徴を用意する必要があった。しかしながら本技術は、LLMが持つ自然言語処理能力を活用することにより、個々のユースケースやシナリオに依らず、汎用的かつ高精度にフィッシングサイトを判定することに成功した。

社会情報研究所がこれまでに創出してきた悪性サイトの検知技術はNTTグループのセキュリティサービスや自社防衛で活用されている。本技術を同様に活用するために、さらなる判定精度の向上や判定処理の自動化・効率化を進めていく予定である。

今後は、AIを活用したセキュリティ対策に加え、AI自体のセキュリティ対策等の技術創出に取り組むことで、新たな技術を安心して利用できるようにし、社会をより良くすることに貢献していきたい。

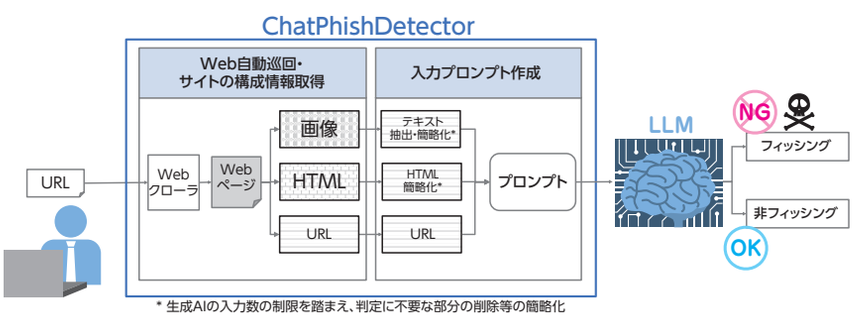

図4　ChatPhishDetectorの処理概要

多様なAIが相互に議論するAIコンステレーション

2019年頃から登場した大規模言語モデル（LLM）は飛躍的な進展を遂げており、汎用的な能力を獲得したAIとしての地位を確立しようとしている。特にChatGPTは、一般にも広く利用され、事業活動への応用も始まってきている。

現状のLLMは、AI開発者のドメイン知識や価値観に基づいてモデル設計を行い、大規模なデータセットから網羅的に知識を学習させることで、ユーザからの自然言語による問い合わせに一定の論理的な回答を実現している。しかし、学習させるデータの網羅性から、得られる回答は一般的な意見になりがちであったり、AI開発者のモデル設計に対する視点に依存したりするなど、AIモデルが多様な価値観で総合的に考えて回答することは困難であるとい

う問題がある。また、学習過程がブラックボックスであり、結果の理由説明や保証ができないことから、意思決定等の高次な判断に利用することが困難であるという問題がある。

・多数のAI連携によるAIコンステレーション

NTTコンピュータ＆データサイエンス研究所では、これらの問題を解決するため、タスクに応じて必要なAIが連携し、集合体として動作する「AIコンステレーション」の実現に取り組む。AIコンステレーションでは、多様な異種AIモデルやルールを環境として与えることで、複数AIが自律的に連携して物理的・論理的世界を広く・深く理解し、人が要因を推測することすら困難な問題に対して多様な視点から解を見出すことが期待される。

例えば、個人レベルでは、仕事とキャリアの変化に適応した健康管理や教育・生涯学習での利益享受が可能になるなど、生活の利便性が向上すると期待される。また、街・共同体レベルでは、交通やモビリティが改善され、地域の安全性や教育アクセスが向上して地域活性化につながるほか、社会全体においては、自動化や連携により産業が変革され、医療やエネルギー分野での進化が期待される。

・AIコンステレーションの実現に向けて

AIコンステレーションの実現には、以下の技術的課題があると考えている。

①AI-AIの協調による議論の高度化

様々な専門性を持った異種AIがそれぞれの視点で意見を表明し、相互に議論・訂正を行うことで、多様な価値観に基づいて複雑かつ大規模な問題を解決する必要がある。特に、各AIの専門性を活かすことで、少数意見についてもその将来価値から互いに重要性を議論し、連携したAI全体の議論を高度化することが必要である。このような異種AI間での協調や議論を高度化するスキームの確立が課題である。

②膨大なAI連携を支える計算機環境

LLMなど大規模なAIの登場により、AIモデルの学習や推論に対するコスト・電力消費量が膨大化している。AIコンステレーションでは、各組織やドメインなどに応じた大規模なAIモデルの学習・推論を行うとともに、広域に分散したAI同士での膨大な知識交換を可能にする必要がある。このようなAI連携を低コストかつ低環境負荷で支えるため、現在取り組みを進めている学習・推論コストを抜本的に改善するAIアルゴリズムの研究開発に加えて、IOWNのネットワーク・コンピューティング環境と融合したアーキテクチャの確立が課題である。

③人-AIの協調による価値創造

AI同士の議論過程やその結論が人や社会に受容されるのかどうか、人から解釈可能である必要がある。特に、複雑かつ大規模な問題になるにつれ、その理解は非常に困難になると想定される。このため、異種AI同士の議論過程を論理的・理論的に解釈でき、なおかつ制御可能にすることで、人とAIが協調できるアーキテクチャを実現することが課題である。

これらの課題を解決していくことにより、複雑化・大規模化した社会問題の解決や新たな価値創造に貢献していく。

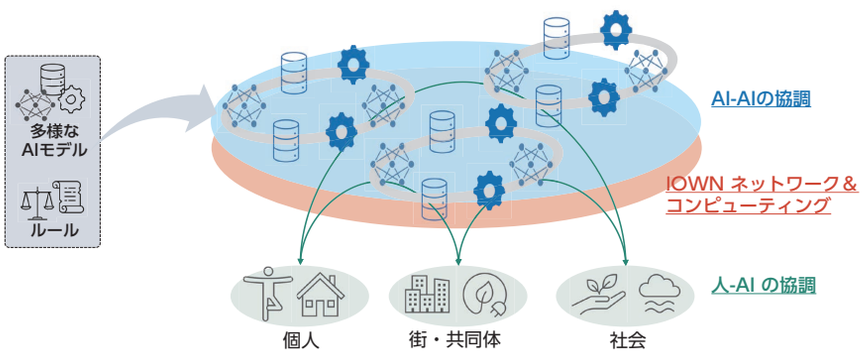

図5　AIコンステレーション

おわりに

サービスイノベーション総合研究所は”ひと”とAIが協働する未来を創造するために、一歩ずつ歩みを進めている。

今回はその大きな一歩“LLM技術”を紹介した。

人類が手に入れた便利な道具AIは、今後、さらに進化し、

高度な遠隔ロボットや、自動システムへと変化し、近い将来様々な形で便利な未来社会を形作ることになるだろう。

我々は、便利な道具AIの使い方を誤ることなく、すべての“ひと”にとってWell-beingな”ひと”とAIが協働する未来を創造していきたい。