

GPTによるレポート自動作成機能を持つ、Web サーベイ AI「ロボリサ」を提供開始

誤情報を検知・削除し、レポートの信頼性を向上

株式会社三菱総合研究所(代表取締役社長:藪田健二、以下 MRI)は、Web からの情報収集および整理・レポート生成を自動化する AI ツール「ロボリサ」に、ChatGPT 等で利用されている大規模言語モデル(以下 LLM)を適用し、高品質な調査レポートの自動生成を実現、提供を開始しました。LLM は誤情報が出力されることが喫緊の課題ですが、「ロボリサ」では誤情報を検知・削除する機能を搭載し、レポートの信頼性を高めています。

1. 背景・経緯

昨今、経営・事業に関するニュースの周辺情報のレポートや、業界動向・研究動向等の調査・整理は、日常的に行われています。ChatGPT に代表される LLM が次々に登場してレポートや情報整理の品質は大きく向上し、これらの業務の効率化が見込まれます。一方で、LLM の出力には創作された内容や不正確な情報が含まれること(誤情報)が大きな問題となっています。

MRI でも、日々レポートや情報整理業務を遂行しています。これら業務の自動化を目指し、MRI がシンクタンク・コンサルティング業務の経験と自然言語処理技術を活かして開発したのが、Web 情報収集・整理・レポート生成ツール「ロボリサ」です。MRI においてロボリサを活用したケースでは、Web 情報収集時間の 80%削減と研究員の知見底上げに貢献しました。

今般、このロボリサに誤情報を検知・削除する機能を実装し、信頼性の高いレポートの生成に成功しました。

2. 概要・特徴

ロボリサは、公開された Web サイトであらかじめ情報収集対象として設定したものの中から日々新着情報を収集、要約し、外国語情報は翻訳した上でデータベースに蓄積します。蓄積した情報は検索・閲覧やメール配信ができるほか、時系列のサマリーを生成するといったレポート機能も有しています。

今回、以下 2 点を実施し、レポート機能を大幅に強化しました(図 1)。

図 1 ロボリサで生成したレポートの一例



出所:三菱総合研究所 文章は GPT-3.5、画像は DALL-E 2 で作成

(a) GPT-3.5 および GPT-4 を用いたレポート自動作成機能の実装

レポートの素材となる文章生成エンジンとして代表的な LLM である GPT-3.5 および GPT-4 を採用し、レポート作成機能を大幅に強化しました。利用者がレポート化したい内容を質問すると、ロボリサは自身のデータベースに蓄積した情報を参考にしてひな型に文章を流し込み、PowerPoint 形式のレポートを生成します。また、画像生成 AI により、レポート内容に沿った画像も生成・挿入します。

(b) 生成された文章に含まれる誤情報を検知・削除する機能の検証と実装

レポートに誤情報が含まれると誤った意思決定を引き起こす可能性があり、望ましくありません。そこで、誤情報が出力される原因を「LLM の処理」と「情報源」の 2 つに分解し、それぞれ対処することで、誤情報を検知・低減できることを検証しました。この結果に基づき、ロボリサは、誤情報である可能性が低い文章のみをレポートに含めます。

3. 今後の予定

LLM により、少量かつ未整理のデータから飛躍的に AI を構築できるようになることから、AI の活用分野は一層広がると考えられます。一方で、誤情報を出力する可能性は残るため、AI の出力を監視する役割の一つである誤情報検知機能は重要です。

MRI は、企業や官公庁の企画業務への DX 導入に積極的に取り組んでいますが、実用的な調査レポートの自動作成が可能になれば、経営企画や研究企画・新製品企画など、情報収集を日常的に実施しているお客さま自身の業務を大幅に軽減できます。MRI は今後、このようなお客さまを中心に、ロボリサの提供を拡大します。そして、法令順守・情報保護・信頼性確保等に引き続き留意しながら、先進的な AI を活用したソリューション・コンサルティングを提供していきます。

本件に関するお問い合わせ先

株式会社三菱総合研究所
〒100-8141 東京都千代田区永田町二丁目 10 番 3 号

【サービスに関するお問い合わせ】

DX 技術本部 高橋、松田、清水
電話:03-6858-3694 メール:robot_research@ml.mri.co.jp

【報道機関からのお問い合わせ】

広報部
メール:media@mri.co.jp

本資料は、経済産業記者会、経済産業省ペンクラブおよび当社にてコンタクトのある記者の方々にもご案内しています。

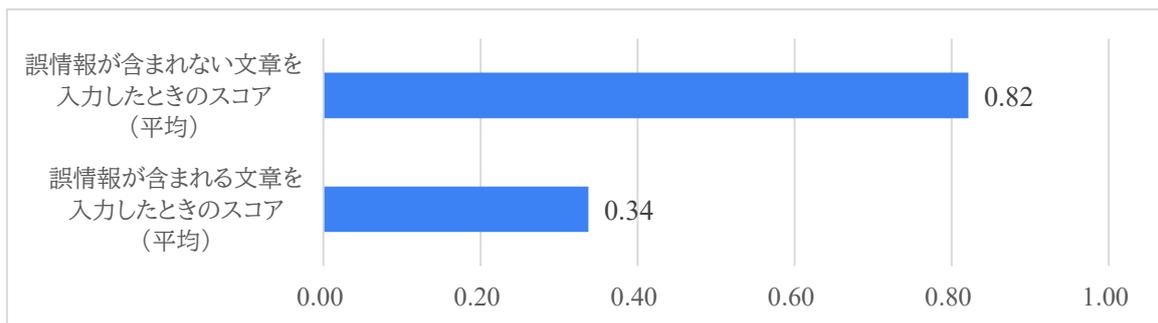
【参考資料】 誤情報の検知・削除に関する検証詳細

① LLM の処理過程で生成される誤情報の検知

AI が生成した文章の評価手法はすでに研究が重ねられており、条件によっては良い評価が可能なが知られている。今回は、UniEval¹という文章の一貫性や流暢さを評価する手法を用いて誤情報の検知を試行した。具体的には、ChatGPT と同世代の LLM²により生成された文章と、文章生成の際に参照した記事との一貫性を評価し、スコアが低い場合は誤情報が含まれる可能性が高いと判定した。

その結果、誤情報が含まれる場合は大幅にスコアが低く、本手法の有効性が確認された(図 1、表 1)。なお、今回は ChatGPT と同世代の LLM の出力を判定したが、他の LLM でも同様の結果が得られることを確認している。

図 2 誤情報有無による一貫性スコアの違い



出所:三菱総合研究所

表 1 LLM (OpenAI API)で生成した文章の正誤検証結果のサンプル

LLM への指示文 (プロンプト)	LLM からの回答	回答の正誤	スコア
宇宙ゴミに関してどのような対策が行われているか?	スペースデブリ対策としては、スペースデブリを軽減・修復する「スペースデブリ管理(SDM)」の取り組みがあります。	正。参照した文章の内容と一致している。	0.975
日本人初の宇宙飛行士は誰か?	日本人初の宇宙飛行士は、1992年にスペースシャトル・エンデバーに搭乗した毛利衛さんです。	誤。参照した文章の内容と一致していない(日本人初の宇宙飛行士は毛利衛氏ではない)。	0.142

② 信頼性の高い情報源の利用

ロボリサの情報源は利用者があらかじめ登録したサイトであり、情報の信頼性は高いと考えられる。一方、LLM の学習で利用したり LLM が回答作成時に参考にするサイトには信頼性が低いものも含まれる。そこで、情報源をロボリサにより取得した記事にした場合と、web 検索により得られた記事にした場合を比較。正確な情報源を使うことにより、品質の高い結果が出ることを確認した。

¹ M. Zhong et al, "Towards a Unified Multi-Dimensional Evaluator for Text Generation," arXiv, 2022. <https://arxiv.org/abs/2210.07197>

² text-davinci-003 および gpt-3.5-turbo