

大規模言語モデルを用いた我が国の対米外交における 調書作成支援システム

原田武夫¹

¹株式会社原田武夫国際戦略情報研究所
takeo.harada@haradatakeo.com

概要

少子高齢化の続く我が国においては知識伝播と技能伝承の刷新が不可欠である。本研究では外務省において生じているそうした問題状況への対処を図るべく、大規模言語モデル (LLM) と検索拡張生成システム (RAG) を用いた業務支援システムを提案する。具体的には日米外交を念頭に、データセットとして党派色の無い米連邦議会調査局 (CRS) によるアジア関連レポートの英文テキストを用いてコーパスを作成し、その調書を日英文で作成、両者の比較を通じての改善案作成を行った。その際、日英文で同一内容の質問を行う中、模範となるべき英文での回答の生成に際し生じる深刻な「幻覚」により、実装に際しての負担増が生じ、課題となることが判明した。

1 はじめに

我が国においては少子高齢化が加速度的に進みつつあり、官民を問わず組織の将来を担うべき若い世代の構成員に対して、如何に効率的な形で知識伝播と技能伝承を行うべきなのかが重大かつ喫緊な課題であり続けている。そうした中で中央省庁もその例外ではなく、「グローバル化」「デジタル化」「金融資本主義化」といった時代の趨勢に伴う行政需要の増大と複雑化によって、そこでの問題状況は深刻化の一途を辿りつつある。とりわけ我が国とグローバル社会の接点にあって外交をつかさどる外務省においてはこの傾向が顕著であり、問題の解決のため、外務省自身が人工知能の実装も含め、改善を検討してきた経緯がある[1]。

状況をさらに複雑にしている要因の一つとしてここで指摘しなければならないのが、米国における第二次トランプ政権の成立である。これまで我が国の外交は「日米同盟」を基軸として展開されてきたが、友好国に対しても応分の負担を求めるドナルド・トランプが大統領職に返り咲いたことで、こうした戦

後の日本外交における大前提が崩れつつある。時に偏った発言を行うことで相手国に揺さぶりをかける手法をとる同大統領の下にあって、米国が果たして如何なる針路を辿っていくのかを探るためには、中立的な立場から米国として取り得る政策オプションを示す公的文書の分析が益々重要になっている。

以上の問題意識に基づき本研究は我が国の対米外交における実務支援システムを提案する。具体的には政策の企画立案に際し、我が国外務省で行われる典型的な業務である調書作成の支援システムを構築する。その際、(1) まずは外交実務支援における人工知能の活用に関する先行研究について述べた上で、(2) 本件支援システムにおいて採用する手法として大規模言語モデルに検索拡張生成システムを付加して用いることを明らかにし、同時にデータセットとしては米連邦議会調査局 (US Congressional Research Service) が同議員らのために作成しているレポートのテキストデータを利用することを述べた後、(3) 「日米外交」とその周辺について検索を行い、調書作成の骨子作成を行う実験を実施し、これに対する考察と、とりわけ外交の現場におけるミドルマネジャーの負担軽減という観点から評価を行いつつ、今後に向けた課題を述べることにしたい。

2 関連研究

人工知能と外交に関する研究の前駆的段階というべき研究として「デジタル外交 (digital diplomacy)」に関する研究がある。この分野における研究の典型例としては[2][3][4]があるが、そこでの関心は主にパブリック・ディプロマシー (public diplomacy) におけるデジタル手段の利活用にあり、本研究が対象としている外交実務の効率的な実施とそこにおける知識伝播と技能伝承には向けられていない。

また、外交実務における人工知能の活用をそもそも社会課題として認識するかどうかは、研究の行われる当該国の経済発展度に伴う人口動態とも深く関

係していると考えられる。例えば[5]は生成 AI を中心とした人工知能の外交における活用について包括的な研究を行うものであるが、外交実務経験を持つ者を含めた執筆者はいずれも開発途上国の出身であることから、こうした問題意識に言及することはなく、外交実務における「ルーティンタスクの自動化 (automating routine tasks)」について言及するに止まっている。

他方、我が国と同じく少子高齢化の波が押し寄せている先進国における研究においても、外交実務における業務支援システムにつき、そこで利用するアルゴリズムを含め、詳細に論じるものは存在しない。[6]はドイツ外務省における大使経験者を共著者を含む研究であり、外交交渉における人工知能の活用可能性について論じるものである。しかし、具体的にどの様なアルゴリズムを用いた支援システムを構築すべきかについては言及することがなく、あくまでも外交における人工知能の活用可能性について触れるに止まっている。

我が国国内における研究状況を鑑みても、本研究の問題意識である「少子高齢化の中における外交実務の効率的な執行、並びに知識伝播と技能伝承のための人工知能を活用した業務支援システムの提案」に関する先行研究は存在しない。

3 採用する手法

本研究が提案する外交実務支援システムは具体的にいうと、我が国の対米外交における調書作成を念頭に置いたものである。調書とは特定のテーマについて収集した情報についてとりまとめ、政策決定の際の参考資料として付する文書を指す。そこで対象となる情報が収録されている媒体の形態は様々であるが、ここでは次章で詳述するとおり、米国の中央政界において広く利用されている文書上のテキストデータを用いる。そこでいう文書は時宜に応じて多種多様であることから、都度コストをかけて大規模言語モデルをファインチューニングすることは迅速性を旨とする外交実務上、現実的では無い。したがって大規模言語モデルに対し、検索拡張生成システム[7]を実装して用いることで本件支援システムを構築することとする。

また本件支援システムは当該業務を効率的に行うと共に、これを使用することが想定される知見・経験共に乏しい末端担当職員が作成する調書の骨子原案に対してチェックをする中間管理職にかかるコス

トも有意に削減するものでなければならない。そのため日本語で作成される骨子原案に対して、それとは別の模範例 (英文) も作成し、後者をベースに前者を数値の段階で明示的に評価した上で、改善案を提示するところまで出来ることがこのシステムには求められる。図 1 は以上を踏まえた本件業務支援システムのワークフローを示したものである。

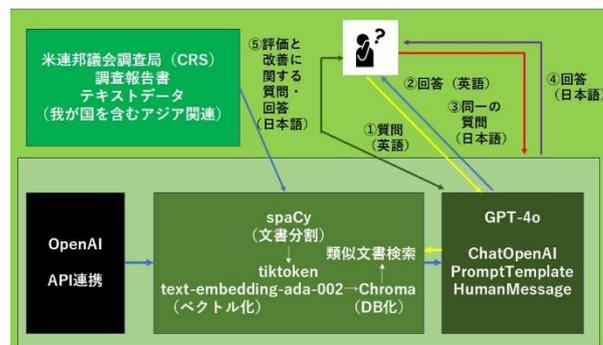


図 1 我が国の対米外交調書作成支援システムのワークフロー

具体的にはまず google colabatory 上で Python 3.9.13 を用いて環境構築を行い、LangChain を利用した。その上で次章にて詳述するコーパスを SpacyTextSplitter により chunk_size は 2000、chunk_overlap は 200 に指定しつつ、分割する。次に API 連携により Open AI の text-embedding-ada-002 を用いてベクトル化した。その際、tiktoken も用いた。ベクトル化したデータは Chroma を用いてデータベースに格納し、質問に対してデータベースから類似度の高い文章を検索出来るようにした。さらに ChatOpenAI, PromptTemplate 及び HumanMessage をインポートし、質問とそれに対する上述の検索結果を組み合わせる形でプロンプトを作成した。大規模言語モデルについては[8]が指摘するように、公開情報上、我が国政府当局への接触の深さを念頭に置きつつ、Open AI の提供する GPT-4o を利用することとし、上述のプロンプトによりこれ呼び出す形とした。

次に、まずは簡単な英文で質問を行い、同じく英文で生成文 1 を作成させる。次に同じ内容の質問を日本語で行い、日本語で生成文 2 を作成させた。その上で Open AI 上のセッションを改め、(1) 生成文 1 を日本語へ翻訳したものをベースとしながら生成文 2 を 10 段階評価させ、問題点を指摘させながら、(2) 生成文 2 に対する改善案を明示的に提示する様に指示を行った。(3) これによって最終的な調書案である生成文 3 が出力される。

4 データセット

本研究が提案する外交実務支援システムは対米外交における調書作成を対象としたものである。米国における政治・外交は民主・共和二大政党の党派色が濃く反映されている。そうした中で党派色の強い政策はメディアを通じて喧伝されがちだが、その結果、背後において実質的な問題の核心をややもすれば隠蔽してしまう危険性がある。したがって我が国の外交政策の立案に際してはそうした党派色を超えて米国が何を利益としているのかを客観的につきとめる必要があり、その意味で超党派(bipartisan)の立場から当該イシューについて米国の見解を述べたテキストをここでは用いるのが妥当と判断した。

具体的には米連邦議会調査局が所属議員とそのスタッフのために刊行している、時宜に応じたイシュー毎の報告書のテキストデータを利用する。その際、今回は便宜的に、同報告書を見やすい形でリストアップした米国科学者連盟 (Federation of American Scientists) の公式 HP[9]に掲載されているデータ中、外交政策・地域情勢 (Foreign Policy and Regional Affairs) 内の我が国を含むアジア (Asia) の項目にアップされている報告書データ (全 263 本。記載型式は様ではない。) をまずは PDF にてダウンロードした。そしてこれを Word ファイル経由でテキストファイルへ転換、全結合し、コーパスとした。

5 評価手法

外交実務上、調書作成に際しては末端担当職員が原案を起案し、これを上長が決裁を行う。そのため本件業務支援システムの生成文に対する評価も、

(1) 調書の骨子原案としての品質に対する評価と、(2) 原案に対する決裁に際し上長にかかる負担がどれだけ軽減したのかについての評価の 2 つがある点に留意する必要がある。この内、前者については英文で作成した生成文 1 (図 1②) をもって和文の生成文 2 (同④) を本件システム自体に 10 段階で自動性能評価させ、かつその理由を明記させた。その上でそこでの評価結果と改善案として創られる生成文 3 を合わせて人手による評価 (human evaluation) を行った。その際、筆者は外交実務上、決裁権者であった経験を持つことから、現場目線を踏まえつつ、この任にあたった[10]。

6 実験計画と結果

本研究で提案する外交実務支援システムが実際に用いられる状況として、まずはより知見・経験に長けた上長がデータセット中、該当する報告書を質問に応じつつ英語原文で適切にピックアップし、それを英文に要約、次に下僚である末端担当職員が別途、同じ質問に対し報告書をピックアップし、日本語で要約したものを、前者の要約 (英文) をベースに評価・改善するというワークフローを想定した。具体的には次のとおりプロンプトを入力した。

・英語で「回答者は日本国外務省の北米担当であること」「その責務はアップロードされたデータベースの中から日米同盟及びその重要性、並びに日本の経済政策に関する米連邦議会調査局作成の報告書を要約することである」「的確に要約は作成し、参照した報告書のタイトル等を記すこと」「アップロードした 2021 年までの報告書だけを参照すること」と指示した (図 1①)。これに対して英語で回答を得た (同②。「生成文 1」)。

・次に日本語で全く同様のプロンプトを入力し (図 2③)、これに対して回答 (生成文) を得た (同④。「生成文 2」)。

・さらにいったんセッションを改めた上で「あなたは有能な AI アシスタントである」「生成文 1 をベースに生成文 2 について 10 段階評価を下し、その理由を明記すること」「生成文 2 を日本語訳すること」「その上で生成文 2 を生成文 1 に基づきつつ改善すること」とのプロンプトを入力 (同⑤)。最終的に回答を得た (同⑥。「生成文 3」)。

As an officer in charge of US political affairs in the Japanese Ministry of Foreign Affairs, I have reviewed the Congressional Research Service (CRS) documents available until 2021 to assess the future of the US-Japan alliance and its implications, as well as Japan's policies in finance and economy. Below is a summary of relevant reports:

1. **“The U.S.-Japan Alliance: Issues for Congress” (July 2021)**:** This report discusses the strategic importance of the US-Japan alliance in maintaining regional stability in Asia. It highlights the alliance's role in countering regional threats and promoting security cooperation. The report emphasizes the need for both nations to adapt to evolving geopolitical dynamics, including China's rising influence.

2. **“Japan-U.S. Relations: Issues for Congress” (March 2021)**:** This document provides an overview of the bilateral relations between the US and Japan, focusing on economic and security cooperation. It underscores the importance of the alliance in supporting Japan's defense capabilities and outlines potential areas for further collaboration, such as cybersecurity and space exploration.

3. **“Japan's Economy: Challenges and Policy Responses” (February 2021)**:** This report examines Japan's economic challenges, including demographic shifts, slow GDP growth, and fiscal sustainability. It outlines Japan's policy responses, such as structural reforms and monetary policy adjustments, aimed at revitalizing the economy. The report also discusses the implications of these policies for US-Japan economic relations.

4. **“U.S.-Japan Trade Relations” (October 2020)**:** This report analyzes the trade dynamics between the US and Japan, highlighting key issues such as tariffs, trade imbalances, and market access. It discusses ongoing trade negotiations and agreements, emphasizing the significance of a robust trade relationship for both economies.

These reports collectively underscore the strategic and economic significance of the US-Japan alliance and highlight areas for continued cooperation and policy development. They provide valuable insights into how both countries can strengthen their partnership to address regional and global challenges effectively.

図 2 生成文 1

申し訳ありませんが、指定されたデータベースや文書に直接アクセスすることはできません。ただし、提供された情報をもとに一般的な知識に基づいて回答を作成することはできます。以下は、指定された情報をもとにした要約の例です。

(1) 日米同盟の未来とその重要性について:

参照レポート: "The U.S.-Japan Alliance" (更新日: 2019年6月13日)

このレポートによれば、日米同盟はアジアにおけるアメリカの安全保障の要であり、地域の平和と繁栄に貢献しているとされています。また、提供された情報の一部は、ユーザーが求めている「日米同盟の未来とその重要性」や「金融及び経済の分野における日本の政策」に関する情報としては不足しているため、評価は低めです。

(2) 金融及び経済の分野における日本の政策について:

参照レポート: "Japan's Currency Intervention: Policy Issues" (発行年: 不明)

このレポートは、日米経済関係の重要性と現状、そして関係管理のための可能な選択肢について探求しています。具体的には、日本の通貨介入政策が取り上げられており、アメリカの政策立案者がこの問題をどのように扱うべきかについての議論がなされています。特に、113回連邦議会が日米経済関係に関する責任を果たすのを助けるための情報が含まれています。

これらは、一般的な知識に基づいて提供された情報をもとにした要約です。正確な情報を得るためには、指定されたレポートを直接確認することをお勧めします。

図3 生成文2

[4]
[1]**アシスタントの回答の評価**:
アシスタントの回答は、指定された要求に対応していません。具体的には、依頼された文書の参照タイトルや発行年の情報が欠落しています。また、提供された情報の一部は、ユーザーが求めている「日米同盟の未来とその重要性」や「金融及び経済の分野における日本の政策」に関する情報としては不足しているため、評価は低めです。
[2]**英語回答の日本語訳**:
以下が提供された英語回答の日本語訳です。
アメリカの政治に関する業務を担当する日本の外務省の担当官として、2021年までに入手可能な米連邦議会調査局(CRS)の文書を精査し、日米同盟の未来とその影響、ならびに日本の金融と経済政策について評価いたしました。以下に関連する報告書の概要を示します:
1.**「日米同盟: 議会への課題」(2021年7月)**: この報告書は、アジアの地域安定を維持するための日米同盟の戦略的重要性について論じています。同盟の強さを維持し、安全保障を促進する役割を強調しています。この報告書は、中国の影響力の増大を含む、地政学的なダイナミクスに対応するため、同盟が適応する必要性を強調しています。
2.**「日米関係: 議会への課題」(2021年3月)**: この文書は、アメリカと日本の二国間関係の概要を提供し、経済と安全保障の協力に焦点を当てています。日本の防衛能力を支える同盟の重要性を強調し、サイバーセキュリティや宇宙探査など、さらなる協力のための潜在的な分野を概説しています。
3.**「日本の経済: 課題と政策対応」(2021年2月)**: この報告書は、日本の経済的課題、人口動態の変化、低成長のGDP、財政の持続可能性を検討しています。政策対応を概説し、構造改革や金融政策調整などの施策を通じて経済を復活させようとしています。また、これらの政策が日米経済関係に与える影響も議論しています。
4.**「日米貿易関係」(2020年10月)**: この報告書は、アメリカと日本の貿易動向を分析し、関税、貿易不均衡、市場アクセスに関する主要な問題を強調しています。進行中の貿易交渉と合意を議論し、両国にとっての強固な貿易関係の重要性を強調しています。
これらの報告書は、日米同盟の戦略的および経済的意義を共に強調し、継続的な協力と政策開発のための領域を浮き彫りにしています。同盟が地域および世界的な挑戦に効果的に対処するために、どのようにしてパートナーシップを強化し得るかについて重要な洞察を提供しています。

(4)**日本語回答の改善案**:
以下は、日本語の回答に英語回答日本語訳の要素を加え、より包括的なものにしたものです。
[1] 日米同盟の未来とその重要性について:
参照レポート:
- 「日米同盟: 議会への課題」(2021年7月)
「日米関係: 議会への課題」(2021年3月)
「日米同盟: 議会への課題」(2021年7月)の報告書では、日米同盟はアジアの地域安定を維持するための戦略的重要性を持つとされています。特に中国の影響力の増大に対応するために、同盟がどのように協力するかが強調されています。さらに、「日米関係: 議会への課題」(2021年3月)においても、経済および安全保障の協力の重要性が強調されています。
[2] 金融及び経済の分野における日本の政策について:
参照レポート:
- 「日本の経済: 課題と政策対応」(2021年2月)
- 「日米貿易関係」(2020年10月)
「日本の経済: 課題と政策対応」(2021年2月)の報告書では、日本が直面する経済的課題とその政策対応について探求しています。特に人口の減少やGDP成長の遅さ、財政持続性について触れ、これらに対する構造改革や金融政策の調整が必要であると述べられています。また、「日米貿易関係」(2020年10月)では、両国の貿易関係、関税、貿易不均衡の問題に焦点を当て、進行中の交渉がどのように影響を与えるかが説明されています。同盟関係の強固な貿易関係の重要性が強調されています。
これらにより、日米同盟と日本の経済政策が持つ包括的な文脈を理解するための情報を提供します。

図4 生成文3

7 考察

図2~4を一見する限り、本件支援システムのワークフローは円滑に機能しているかの様に見える。しかし精査の結果、生成文1で列挙されている4つの「報告書」は今回利用したコーパスを収容したデータベースに一切含まれていないことが判明した。さらに米連邦議会調査局の公式HP[11]で検索しても、生成文1に明記されている4つのタイトルで公表された「報告書」は実在しないことを確認出来た。すなわち生成文1は完全なる「幻覚(hallucination)」であったことになる。これに対して生成文2で取り上げられている2つの「報告書」は上記データベースに含まれているものである。しかし我が国に関する「報告書」は今回19本含まれている中、2本のみを抽出している点が課題として残された。さらに生成文3について見ると、生成文2についての10段階評価を行い、その理由について述べている部分は有効に機能していた。ただし10段階評価の基準に関する詳細については説

明が付されていなかった。それ以上に改善案の提示に際し、生成文2に盛り込まれている要素を完全に抹消した上で、「幻覚」に過ぎない生成文1の日本語訳をベースに再編集したものを提示していた点が特筆に値する(ただし、その際の項目立てについては質問文に即した形で的確に記していた)。

以上を踏まえて人手による評価を行うならば、まず末端担当職員の観点から効率的な作業を本件支援システムで行えるのが基準となる。これについては日本語での質問に対する生成文が内容的に乏しく、また英語で同趣旨の質問をすると「幻覚」が生じてしまう点で効率性の向上に資するとは考えにくい。次に末端担当職員の作成する調書の骨子原案を精査する管理職の立場からすれば、10段階評価の実施やそれに対する理由付け、また生成文1をベースとした改善案提示はある意味、コスト削減につながるかと考える。しかし生成文1がそもそも「幻覚」であり、かつ改善案については真正な「報告書」をスキップし、「幻覚」のみに依拠して作成していることから、最終的にはむしろコスト増をもたらしている。以上の現象は複数回にわたり試行を繰り返しても結果は同じであった。

8 総括と今後の課題

同一の英文コーパスに対して日英で同義の質問を行うのに対して、(1)和文については検索拡張生成システムで類似度の高い文章が引き出されるのに対し、英文では「幻覚」で回答が生成されるのみならず、(2)後者で前者を評価・改善せよとの指示を出すときも後者についての最終的な改善案が生成されるという本研究の結果は、提案する外交実務支援システムが実装された場合、外交上、深刻な影響があり得ることを示唆している。外務省に限らず、我が国政府機関で利用すべき大規模言語モデルの採用に際してはこのことを踏まえ、(政策立案に対する外部からの操作可能性の観点からも)慎重な考慮が必要であることが確認出来た。図1で示す支援システムのワークフローについて、(1)データセットの前処理をmetadataなどの利用により更に精緻に行うことで類似文の抽出をより広範囲に行う、(2)生成文を外交実務上の調書として用いるためにはより長文を生成させる必要がある、といった今後に向けた技術的な課題はあるものの、前述の論点はより根本的なものであり、実装にあたっては第一に注視が必要と考える。

参考文献

1. 大規模言語モデルの我が国外交における活用: 外交文書の要約業務における実装に関する考察. 原田武夫. 学習院女子大学紀要, 2024 年, 第 26 巻. 225-265.
2. Social media and public diplomacy: A comparative analysis of the digital diplomatic strategies of the EU, US and Japan in China. **Bjola, Corneliu, Lu Jiang**. Digital diplomacy. Routledge, 2015, 71-88.
3. Coping with Digital Disinformation in Multilateral Contexts: The Case of the UN Global Compact for Migration. **Bjola, Corneliu**. Digital Diplomacy and International Organisations. Routledge, 2020, 267-286.
4. Digital diplomacy in the time of the coronavirus pandemic: Lessons and recommendations. **Bjola, Corneliu, Michaela Coplen**. The Palgrave handbook of diplomatic reform and innovation. Cham: Springer International Publishing, 2023, 323-342.
5. The Role of Generative AI in Global Diplomatic Practices: A Strategic Framework. **Bano Muneera, Zahid Chaudhri, Didar Zowghi**. arXiv preprint arXiv:2401.05415.2023.
6. Diplomacy and artificial intelligence: reflections on practical assistance for diplomatic negotiations. **Stanzel Volker, Daniel Voelsen**. 2022: 31.
7. Retrieval-augmented generation for knowledge-intensive nlp tasks. **Lewis Patrick, Perez Ethan, Piktus Aleksandra, Petroni Fabio, Karpukhin Vladimir, Goyal Naman, Küttler Heinrich, Lewis Mike, Yih Wen-tau, Rocktäschel Tim, Riedel Sebastian, Kiela Douwe**. Advances in Neural Information Processing Systems, 2020 年, 第 3 巻. 9459-9474.
8. 国会答弁案作成システムにおける生成 AI の活用「対北朝鮮外交」を題材とした実装の検討. 原田武夫. 人工知能学会全国大会論文集 第 38 回 (2024). 2024.
9. **Federation of American Scientists**. Congressional Research Service Reports. Foreign Policy and Regional Affairs, Asia: <https://sgp.fas.org/crs/row/index.html>, 2024-11 閲覧.
10. 国会答弁案作成システムにおける生成 AI の活用「対北朝鮮外交」を題材とした実装の検討. 原田武夫. 人工知能学会全国大会論文集 第 38 回 (2024). 2024.
11. **Congressional Research Service** : <https://crsreports.congress.gov/>, 2024-11 閲覧.