

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

Summary

本研究は生成系AI (generative AI) の中でも大規模言語モデル (Large Language Model, LLM) に着目し、我が国外交実務におけるその活用を巡る可能性について考察するものである。我が国では少子高齢化が加速度的に進み、行政実務の効率化がこれまで以上に強く求められてきている。とりわけグローバル化が進み、行政需要が一段と高まっている我が国外交実務の現場においてそうした行政の効率化の一環としてデジタル・トランスフォーメーション (digital transformation, DX) の必要性は高まるばかりであるが、人工知能 (AI)、特に生成系AIをそこで積極的に活用しようという本格的な取り組みは依然行われていない。他方で我が国では官民を挙げて生成系AIの社会実装の拡大に取り組み始めていることは事実であり、こうした流れの中で我が国外交実務においても生成系AIの活用について検討することは不可避であると考えられる。以上を背景としつつ本研究ではOpenAI社が提供するChatGPT-4により、外交実務においてしばしば行われる英文及び和文テキストの要約 (summarization) を行い、BERTScoreによる評価を行った。実験の結果、英文テキストと比べ、和文テキストについて特に長文であるとプロンプト (prompt) を相当程度工夫しなければ要約を作成することは困難であることが判明した。また英文テキストについては比較においてこうした問題は見られないものの、評価指標上、どうしても生成文の修正を必要とする結果とならざるを得なかった。以上を踏まえ、生成系AIの我が国外交実務における活用の際には、その成果物をチェックする管理監督者の立場にある者のタスクが増えることに留意しなければ行政の効率性を即座に高めることにはならないことが明らかとなった。

1. はじめに

我が国では依然として人口減少が加速度的に続いており、仕事の現場では技能伝承と知識伝播が問題であり続けている [樋口ら 21]。このことが民間企業のみならず、行政についても当てはまることは言うまでもない。さらに言うところ行政に含まれる外交についても状況は深刻であることをまずは指摘しておかなければならない。グローバル化が一

層の進展を遂げ、国際社会では複合的なリスクが続々と出現しているものの、我が国外務省の定員は過去10年間で約13パーセントしか増員されていない [外務省 23]。そのため我が国外務省員は少数精鋭とならざるを得ないものの、そこで行われる外交実務はより少ない資源の投入であっても、従来以上に質の高い成果を示さなければならないようになってきているということは [原田 23] において指摘したとおりである。

[原田 23] はさらに、我が国において他の行政機関と同様に外務省に対しても行われている政策評価（「行政機関が行う政策の評価に関する法律」（平成17年12月16日閣議決定）並びに同第6条に基づく「政策評価に関する基本方針」（平成30年2月6日策定・同7月1日改定）による）に関し、「政策評価の観点に関する事項」として同法第6条第2項第3号に基づき定められている事項を取り上げている [外務省 18]。

外務省が政策評価を実施するに当たっては、以下の観点を基本としつつ、評価対象政策の性質等に応じ、適切な観点を選択し、総合的に評価する。

- (1) 必要性：政策の目的が国際社会における日本国及び日本国民の利益の増進という観点から妥当か。行政関与の在り方から見て国がその企画立案及び実施の主体となる必要があるか。
- (2) 有効性：政策の実施により、期待される効果が得られるか、又は実際に得られているか。
- (3) 効率性：投入された資源量に見合った結果が得られるか、又は実際に得られているか、政策効果の発現のためにとられる手段は適切、効率的であるか、など。

本論文では [原田 23] と同じく、ここで掲げられている3つの観点の内、効率性に着目する。我が国の行政機関における効率性を巡る議論としては、これを算出（output）／投入（input）の最大化と定義する [秋吉ら 20] がある。また各国の会計検査ではほぼ共通に使用されている4つの検査基準の一つとして効率性をとらえ、それを前提としながら行政活動の「能率」について考察する研究 [西尾 01] もある。加えて地方自治体における財政運営と透明性、さらには事業効率性について論じた研究として [山下・赤井 05] がある。これに対して政策評価という視点から、特に行政機関としての外務省に着目し、そこでの効率性とその改善のための具体的な手法について考察した論考は存在しない。

本論文は外務省における実務を外交実務と呼んだ上で、その中でもとりわけ非熟練の末端担当職員が日々こなすことを期待されている外交文書の要約（summarization）をより短時間で正確に遂行するという意味での効率性を高めるための方途として大規模言語モデル（Large Language Model, LLM）を活用すべきか否かについて検討することを目的としている。[原田 23] では先行研究において我が国外交実務に対する政策評価

に際し、効率性の向上を対象とした具体的な手段につき、人工知能を活用する研究が行われていないことを確認した。こうした状況は本論文執筆の段階でも管見の限り変わっていない。大規模言語モデルを用いた、この意味での外交実務上の効率性向上を考究した先行研究は存在しない。

本論文の構成は次のとおりである。第2章では関連研究について考察する。具体的には我が国外務省に対する政策評価と、そこでの規準としての効率性を巡る議論全般を巡る先行研究に触れる。次に我が国における外交実務において「要約」が持つ意味合いを取り上げる。その後、外交実務における人工知能の活用について確認をする。第3章では採用する手法として大規模言語モデル、具体的にはOpenAI社の提供するChatGPT-4について概観し、その上で今回用いる外交文書のデータセットについて述べる。第4章では評価方法、第5章では実験計画と結果、第6章ではこれに対する考察を述べる。以上を踏まえ、第7章では総括と今後に向けた課題について述べる。

2. 関連研究

2. 1. 我が国外務省に対する政策評価と規準としての効率性

[原田 23] は我が国外務省における政策評価に関して検討するに際し、[吉原・山谷 09] に依拠しつつ、およそ政策評価が外交実務の現場における実情に対して適正に行われていないことを指摘する。例えば吉原は自らが外務省における政策評価の担当官であったという体験を踏まえつつ、次の様な傾向がそこでは見られると述べている。

- (1) 抽象度が高く難解な外交は、プロフェッショナルな外交官の個人的な力量と長年にわたる努力に成功が支えられているのであって、これを毎年素人に分かりやすく評価して、公表することが可能とは考えられないという声が聞かれる。
- (2) 外交の成果や進捗状況を数字で表しにくい抽象的な価値で示し、政策目標にする場合があり、これを数値化することは困難であるという議論も聞こえてくる。
- (3) 長年努力し続けてもうまくいかなかった懸案課題が我が国政府の関与しない国際環境の変化によって好転する場合も少なくない。

また竹本は我が国外務省に固有の人事政策における硬直性を指摘する [竹本 11]。すなわち (1) 外務省内では各地域を担当するグループが入省後の語学研修制度を契機に形成され、それがその後の省内キャリアパスに結び付いていること、(2) 外務公務員は長期間の在外研修、外交官特権、大使任命の際には皇居にて認証官として天皇から直接認証を受けること、等から他のキャリア国家公務員と比べても特別な処遇を受けていること、を指摘している。そしてこれらの結果、外務公務員たちの間では超然主義的な態度を帯びた「理想主義」の傾向が強まり、外務官僚の「自律性」とでもいうべき根本的

な態度が形成されるとしている。外務公務員たちの間で言うならば「エートス」として受け継がれる自律性を背景としながら、外部からの評価を当然には受けないという姿勢がそもそも形成されてきており、かつこれが硬直的な省内キャリアパスを通じて、外交実務に必要な技能伝承や知識伝播が広く展開するのを阻害しているとするれば、これこそが正に政策評価という効率性の規準を阻むものとして是正されなければならないのである。

2. 2. 我が国外交実務と外交文書の要約 (summarization)

前節で述べたような問題意識に基づき [原田 23] は我が国外務省に対する政策評価においてこれまで十分な精査が行われてこなかった規準である効率性につき、具体的な改善を施す契機は外交文書の起案という、外交実務における基本中の基本について改善を施すことにあると指摘する。行政の効率性を引き上げるためにはその産出 (output) を増やしつつ、それに応じる形で資源投入 (input) を適正化することが不可欠である。こう考える時、外交実務において投入 (input) される資源の内、特に人材の量については我が国における少子化の加速を前提に減少する一方である。それ以外の資源としては「労働時間」と「カネ (予算)」があるが、我が国において恒常化している財政赤字を前提とすれば後者について増やすことは現実的ではない。しかも「ヒト (人員)」については我が国政府部内で総務省行政管理局、「カネ (予算)」については財務省主計局が担当しており、外務省が一存では決められない点に特徴がある。これに対して「労働時間」については外務省自身の努力で削減することは可能なのである。[原田 23] はこの様に考える延長線上で、外交実務における労働時価の削減のためにはそこでの政策プロセスの中核をなす稟議制における文書起案の迅速化が必須であると述べる。そして我が国の外交実務上、文書起案は若手の末端担当職員が行うことが通例である中、未熟練で外交実務の現場での知見に乏しいこれら末端担当職員が如何に素早く、文書起案のための情報を、時として膨大な量に及ぶ過去の外交文書から迅速に得ることが出来るかが、我が国外交実務における効率性を改善するにあたって鍵を握っているとする。

本論文は [原田 23] で提示した以上の様な問題意識を共有しつつも、我が国外交実務における効率性を引き上げるためのアプローチとしてはそこで取り上げたトピック・モデル (Topic Modelling) の活用ではなく、生成系AIの中でも大規模言語モデルを用いた外交文書の「要約」に着目するものである。[原田 23] は我が国外務省において外交文書が起案されるパターンとしては概ね次の2種類があると指摘した。

(パターン1)

上長より末端担当職員が口述を受けて作成する場合

(パターン2)

上長からの指示を受けるが口述は行われず、関連の文書ファイルを参照しつつ末端担当職員が作成する場合

より実務に即して述べるならばいずれの場合においても、稟議制のプロセスに起案した文書をのせる時、大抵の場合に必要となってくるのが「関連資料」及び「参考資料」である。これらは起案した外交文書がどのような契機・理由に基づき起案されたものなのかを起案者より上位に位置する決裁権者に対して端的に示すものである。決裁権者の下には恒常的に大量の決裁案件が上がって来ることから、こうした「関連資料」「参考資料」は本体である(起案されたばかりの)外交文書が「直ちに決裁されなければならない理由」を極力少量の文章で鮮明に示すものでなければならない。特にそれまでの経緯が長い外交事案の場合、関連する文書の量も多く、かつ個別の文書も大量のテキストから構成されている場合がままある。あるいは全く新しい分野での案件についての処理を求める外交文書である場合、その新規性が故にその問題の背景事情や重要性などについて大量の文書を事前に参照しなければ決裁出来ない場合もしばしばあり得る。そしてこうした時に、多くの場合には若手である末端担当職員が作成を求められるのがこれら文書の「要約」(summary)なのである。グローバル化の中で大量の外交文書の起案と決裁が必要とされている中だからこそ、この意味での「要約」の必要性は高まる一方であると言べきである。

国際社会における案件の中でも昨今は地球規模での案件、すなわち特定国についてのみならず、グローバル社会全体が同時に直面する事案が少なくない点に留意する必要がある。例えば2019年後半から全世界を席捲したCOVID-19がその典型である。そしてこうした場合、問題解決に向けて特定の国々だけではなく、様々な国際機関(international organization)が積極的な取り組みを行う場合が少なくない。この場合、国際機関の事務局の側は加盟国政府による議論を促すため、大量の関連文書を作成する場合がままある。英文で時に数百ページ以上にもわたる文書が作成されるわけであるが、これに対する我が国政府としての対応を決めるに際しては対処方針を固めることが必要になる。この対処方針を記した外交文書を末端担当職員が作成し、上長の了解を得るに際しては、その文書が大部であればあるほど、こうした国際機関作成の文書の要約を迅速に執筆することもまた、当該担当職員の重要な任務となる。

しかし昨今のグローバルイシューの多くが、外交や安全保障といった外務省における典型的な領域ではなく、むしろ公衆衛生や人工知能科学といった、外務省職員が通常、門外漢であるはずの領域に属する点をここでは指摘しておかなければならない。無論、これらの政策分野についてそれぞれの事案を所管する関連省庁が存在しており、その意向が我が国政府の移行となることが外交場裏でもままある。だがそうであるからといって、外務省の内部で全く検討がなされないのかということそうではないのであって、多く

の場合において若手である末端担当職員は、これまで見たことも聞いたこともなかった分野における最新のイシューの書かれた大量の文書を読みこなし、要約しなければならないというのが外交実務の現場における実情なのである。こうした外交文書とその周辺の文書の「要約」業務が、グローバルイシューの急拡大によって拍車をかけられた外交実務の肥大化の中で増加の一途を辿っているのであって、この部分における効率化が、外務省に対する政策評価における規準としての効率性の向上をもたらすと考えられる。

それではここで言う「要約作業」は何によって迅速化するのであろうか。第一に言えるのは場数を踏んだ年長の外務公務員に匹敵するくらいの知見・経験である。第二に、国際機関が作成する文書の多くが英語によるものであることから、卓越した英語能力が求められる。その上で、長い文書であってもその構造を的確にとらえ、要旨を簡潔かつ要領よく日本語で執筆することの出来る能力が必須である。しかし外務省における現状はというと、第1章で述べたとおり、専門とする言語に応じた人事ローテーションが組まれており、その硬直性（あるいはその裏側にあるエートスとしての「自律性」）が故に知識伝播と技能伝承が組織内部で阻まれている。さらに言うとも英語以外の言語を専門語とする外務省職員にとって、長い英文テキストを要約することは骨の折れる作業でしかない。また外務公務員の採用時に「英文テキスト（ないしは和文テキスト）に関する要約文の作成能力」は問われていないのであって、日々の実務を通じてこれを習得していくしか手段がないということもまた実態なのだ。

2. 3. 外交実務における人工知能の活用

〔原田 23〕は、我が国の民間セクターにおいて技能伝承・知識伝播の手法につき種々提案され、実際に活用されているものの、外交実務ではそういった角度からの先行研究はないとする。そして我が国以外における研究動向を見ても、外交実務における技能伝承・知識伝播という角度からの研究は存在しないと述べる。他方で〔蒔田 20〕が政策立案に際しての人工知能の活用を提唱し、かつ〔福元 20〕が計量政治学における人工知能を用いた研究の拡大を述べていることを踏まえつつ、〔原田 23〕は我が国外交実務における人工知能の活用を模索する中でトピック・モデルによる過去の外交文書上のトピックの迅速な把握と、それに伴う外交文書起案の迅速化のための仕組みを提唱する。その際、〔原田 23〕は次の様な問題提起を行っている。

「本論文が考察の対象としているのは外交文書の起案において必要な過去の外交事案とその処理に関する知見の効率的な伝搬である。現状の問題はこうした知見がそもそも「暗黙知」に止まっている点にある。これをどの様にして（若手）末端担当職員に対し効率的に伝搬し、かつ外務省という組織全体で可視化し、もって外交文書の起案時に誰でも容易に利用出来るものにしていくことが出来るのか」

〔原田 23〕が我が国外交実務における人工知能の活用の中でもとりわけトピック・モデルに着目しているのは、外交実務における技能伝承・知識伝播に際し、そこで対象となる知識に関する「情報粘着性」〔Von Hippel 94〕、さらにはそれを主たる特徴とする暗黙知としての性格こそ解消されなければならないという認識に基づいている。そして暗黙知を形式知に転換し、組織内での共有・定着を目指すプロセス概念としての「SECIモデル」〔野中・竹内 20〕に言及しつつ、そこでいう暗黙知から形式知への表出化、さらには形式知から形式知への連結化を外交文書におけるトピックの可視化を通じて行うという方向で議論を行っている。

外務省自身がデジタルトランスフォーメーション（DX）に対する姿勢を初めて明確に示したのは2021年7月に公表した外務人事審議会『ポスト・コロナに向けた外交実施体制の抜本的強化：外交・領事実施体制のデジタル・トランスフォーメーション（DX）の一層の促進に向けた勧告』においてである〔外務人事審議会 21〕。この中では外務省職員の労働環境の改善について次のとおり述べられている。

「DXの一層の推進により、全体として職員一人一人の活躍を支える職場環境の一層の改善を図るとともに、必要な超過勤務に対しては、適正な手当を支払うための必要な予算を確保していくべきである」

一見して明らかなおおり、ここでは外交実務に費やされるべき（労働）時間の縮減を求めている。このことは本論文が基礎とする視点、すなわち外務省に対する政策評価における規準としての効率性の向上のための検討を行うべしとの方向性と合致している。それではこうした議論の延長線上で我が国外務省は人工知能（AI）を用いた問題解決を実務上、試みているのであろうか。

この関連で注目すべきなのが「外務省デジタル・ガバメント中長期計画」〔外務省 20〕である。我が国においてはデジタル・ガバメント実行計画（2019年12月20日閣議決定）に基づき、各府省において府省CIO及び副CIOのリーダーシップの下、各府省デジタル・ガバメント中長期計画を改定してきている。外務省においては2018年6月28日に決定され、2020年3月30日に改定されたこの計画の中には次の様な言及がある。

「まず、2020年度においては、AIの活用により、当省内で翻訳業務を行っている課室に対して精度が高い翻訳業務を行えるよう翻訳業務を支援する基本的な機能及び翻訳エンジン（日、英）の整備を行い、翻訳時間・精度の検証を行う。

その後、2021年度以降においては、段階的に機能拡張（英語表現の多案化機能等）、英語以外の多言語（中・仏等）についても導入を検討していく。

省内で、継続・反復的に行われる業務、意思決定や判断に直接関わらない業務（入力・

転載作業等)、長時間を要している事務作業などから優先的に一部業務の自動処理化を実現するため、RPAの導入について取り組む。」

この言及からは外務省内における翻訳業務について人工知能を導入し、これを活用しようとする試みが行われていることが分かる。しかしそれ以上の業務分野において人工知能の導入が企図されているか否かは、この計画だけでは判然としない。とりわけ外交実務の基本というべき文書作成に関連し、その基礎的な補助・支援ツールとして人工知能の活用が行われていることを積極的に示す文書は管見の限り対外公表されていない。

むしろ「ウクライナ戦争」の勃発を受け、外務省が自ら強化する方向へ動いたのが、人工知能を活用したソーシャル・メディア上の偽情報の検索や分析を目的としたシステムの導入である〔荒木 23〕。これは2022年12月16日に閣議決定された国家安全保障戦略を受けたものであるが、同戦略には「偽情報等の拡散を含め、認知領域における情報戦への対応能力を強化する。その観点から、外国による偽情報等に関する情報の集約・分析、対外発信の強化、政府外の機関との連携の強化等のための新たな体制を政府内に整備する」との記述がある。2023年度外務省予算においてはSNS空間におけるモニタリング、情報分析等に必要経費として、新たに2.1億円が計上された他、国際情勢分析能力強化を見据えた人工知能の活用のため1.4億円が計上された。

以上見てきたとおり、外務省自身が職員の労働環境を改善すべく、人工知能の活用を模索し始めていることは事実である。しかしそこで念頭に置かれている業務分野は、本論文が対象とする「外交官の自律性」を前提とする暗黙知の可視化とそれに伴う外交実務の効率性向上に直結するものとは言い難いものである。確かに「ウクライナ戦争」において典型的に見られたハイブリッド戦争に際し、重要な要素となるインターネット上の「偽情報」に対する警戒措置としての人工知能の活用は、外交実務の現場において我が国外交官たちに対して新たな知見をもたらすものではある。しかしそれはデータサイエンスを活用した上での新規の知見というべきなのであって、組織としての我が国外務省における「暗黙知」とは必ずしも関連するものではない点に留意する必要がある。本論文においては我が国外務省に特有とされ、その意味で課題でもある「外交官の自律性」とも直結するそこでの「暗黙知」を「形式知」として、多くの場合には若手である末端担当職員が簡単に利用するためのツールとしての人工知能の活用可能性に関し、外交文書の一部である要約作成に注目しながら引き続き考察していく。

3. 採用する手法とデータセット

本論文では外交実務に対する政策評価に際し、規準である効率性を向上させるべく、特に外交文書の要約作成に着目して人工知能の活用可能性を探るものである。しかし、この観点より外交文書の要約作成に特化した形で人工知能の利用を模索した先行研究は

存在しない。そこで視点を変えて、外交文書と同等レベルと考えられる文書について要約タスクを課し、そこでの人工知能の活用可能性を検討した他分野の文書を対象とする研究に着目してみることにする。

3. 1. 先行研究に関する考察と手法の選択

[相澤 2023] は自然言語処理による学術論文の解析について現状とその展望を述べる。その中で相澤は「(註：学術) 論文の自動要約は歴史が古い研究テーマである」としつつ、1950年代には既に、単語の頻度をカウントして文の重要度を計算し、スコアが高い文を抽出することで要約を自動生成する手法が考案されていたとする。加えて2000年代に入ると論文に特有の文書構造に基づき要約を生成する手法が提案され、論文の特性を踏まえた要約手法の重要性が認識されるようになったと述べている。また、そもそも自然言語処理の多くは今日、大量の文書を使った学習により、文法や意味構造など多様な言語情報をパラメータの重みとして埋め込んだモデルである「事前学習済の言語モデル (pre-trained language model)」を前提としているとも述べ、とりわけ論文テキスト中では専門用語や論文を対象として言語解析を行う場合は、学術分野に特化したモデルを用いることが一般的であるとする。加えて、自動要約には文書の構成要素を抽出して組み合わせる抽出型 (extractive) と、新たに要約文を生成する抽象型 (abstractive) の2つのアプローチがあることも指摘している。そこで本論文でも以上の指摘を踏まえ、(1) 「事前学習済の言語モデル」(すなわち大規模言語モデル) の利用を前提としつつ、(2) 外交実務に特化した形でのモデルにつき前例の有無をあらためて確認した上で、(3) 「抽出型」と「抽象型」の2つのモデルを念頭に置きつつ、考察を進めていくことにする。

2017年にGoogleが機械翻訳のモデルとして発表したのがTransformerであり [Vaswani et al. 2017]、これを採用した最初の大規模言語モデルが2018年にOpenAI社の開発したGPT (Generative Pre-trained Transformer) である [Radford et al. 2018]。その後、文脈化単語埋め込みを計算するTransformerにつき大規模コーパスを用いた学習で事前学習し、次にこのモデルを下流タスクのデータセットを用いて微調整 (fine tuning) するという手法が自然言語処理における標準的な手法になっていく。GPTにおいてはデコーダ構成が採用されているが、これに対してエンコーダ構成のTransformerを採用することで、文脈化するトークンに先行するトークン列と後続するトークン列の双方からの文約を捉えるように改良を施されたのが同じく2018年にGoogleが提案したBERT (Bidirectional Encoder Representations from Transformers) である [Devlin et al. 18]。その後、大規模言語モデルの開発が進められる中でモデルに含まれるパラメータ数と事前学習に使うコーパスの容量が飛躍的に増加してきている [山田ら 23]。具体的に言うとまずパラメータ数について、2018年に発表されたBERTでは34億個、2019年のGPT-2では15億個であったものの、2020年に発表されたGPT-3では1750億個になっ

ている。また、コーパスについてもBERTのコーパスに含まれる単語数が30億語程度であったのに対し、RoBERTa [Liu et al. 19] は300億語、GPT-3は3000億語である。そしてこのような大規模化が行われている理由としては、モデルの規模を大きくすることで性能が比例して改善していくという「スケール則 (scaling laws)」があることが知られている。以上見てきたとおり、大規模言語モデルを巡っては既に様々なモデルがあり、これを我が国の外交実務の中でも要約作成という業務を例にとりて実装する際には、その中でもどのモデルを選択するかという問題がある。

この課題に対する解を導き出す手順として一番簡単であるのが、外交に特化した形で事前学習が施された大規模言語モデルを利用するというやり方である。しかし管見の限り、その様な大規模言語モデルに関する先行研究は存在しない。スコープをより広げて、官僚制に特有な単語・言語表現を事前学習させた試みとしては [Auriemma et al. 23] がある。これはイタリアの行政機関で用いられる「イタリア語」が通常の「イタリア語」とは意味的及び構文上、大きく異なっているという問題意識に基づき行われた研究をとりまとめた論文である。もっとも「官製イタリア語」だけではなく、学習の対象には保険及び銀行分野での言語表現も含むと同研究ではしており、そうした工夫の結果、特に固有表現抽出 (Named Entity Recognition, NER) について有意な成果を導き出すことが出来たと述べている。これに対して [Auriemma et al. 23] は本論文が対象としている「要約」について触れていない。

他方で我が国における大規模言語モデルの活用につき、行政機関の内部で比較的活発な議論が行われているのが特許行政の文脈においてである。例えば [伊藤ら 23] は我が国の特許庁が特許行政事務の高度化と効率化を目的に2016年度から人工知能技術の適用可能性の検討を着実に進めてきているとして、その方向性とこれまでの成果について検討している。具体的にはBERT、ALBERT [Lan et al. 19]、DeBERTa [He et al. 20]、Longformer [Beltagy et al. 20] を取り上げ、各モデルを特許事前学習モデルとして用いた結果につき比較・評価する。その上で次の様に述べている。

「特許関連タスクの精度向上という難題に対し、特許文献という特殊な文構造等を有する文章をもって事前学習したモデルを用いることが有用であるという今後の可能性を示せたものと思う。(中略) 対話型AI・生成AIの一例であるChatGPTは、社会的にインパクトのあるツールではあるものの、特許行政事務での活用に適しているのかどうかについては、更なる検討が必要だと考える。本事業で用いた各事前学習モデルは、プログラムが公開されており、学習済のモデルも配布されている等、実証にかかるコストが比較的安く、各タスクに適したファインチューニングが比較的容易に行える等の特徴を有する。他方で、GPT系列のモデルは、莫大な学習コストによって作成された、内部が非公開のモデルであるため、通常的环境下では、モデルの中身を踏まえた改良は難しく、

タスクに適合するよう改修するには、相応のコストが必要と思われる。また、入力する情報の保護や出力された結果への信頼性、全体の運用等について、未だ議論が十分にされていない。こうした背景を踏まえ、対話型AI・生成AIについては、特許行政事務の効率化等が見込めるのかという観点だけではなく、情報の取り扱いや運用面等も含め、引き続き注視していきたい。」

以上の関連として付言すると、民間セクターにおける研究であるが「特許調査」に際しての大規模言語モデルの活用可能性について研究したのが [安藤 23] である。その中で安藤は「ChatGPTの学習モデルは2021年9月までのデータを基にしていることに注意が必要である」としつつ、GPT-3.5、GPT-4、さらにはBERTなどを用いた特許調査の結果を検討する。そしてGPTは自然言語による質問入力、入力された文章を要約するタスク等は非常に得意なようであり、タスクを細かく分けてどのような指示を出すか、柔軟な発想が問われるとしている。

以上述べてきたところをまとめるならば大規模言語モデルについて、特に我が国の外交実務を巡る単語・言語表現につき事前学習を行ったモデルに基づく先行研究は存在しない。他方で行政全般へとスコープを広げても本論文が目指している「要約」における大規模言語モデルの活用可能性について行われた先行研究は無い。加えてGPT系列のモデルについては事前学習の際に用いられたデータの限定的な性格や、そもそも内部が非公開のモデルであることによる不透明性などを理由に、実務面での取り扱いには後ろ向きな風潮がある様に見受けられる。

こうした中で事態を打開するための策として考えられるのが、fine tuningに際して我が国外交上の単語や言語表現を集中的に組み込み、学ばせるという手法である。ただし、これについては [Hong et al. 22] が14のBERT型の大規模言語モデルに対して他分野にわたり収集された2210億トークンの科学用語をもってfine tuningした結果を分析している点が参考になる。具体的には11のタスクをこれらのモデルによりこなさせる中で、fine tuningを行っても結果につきあまり大差はなかったと結論づけている。したがってこの論文のタイトルにあるとおり（原題 “ScholartBERT: Bigger is Not Always Better”）、大規模言語モデルについてはfine tuningを行うなどしてより規模を大きくしたとしても、一方では確かに上述の「スケール則」が働くものの、他方では目に見えた性能向上を達成出来るとは限らない点に留意しておく必要がある。

またモデルの詳細が非公開であるとの指摘を受けているGPT-4についてであるが、確かにこれについて開発者であるOpenAI社が公表しているのは特定の環境下での性能を記しただけの資料のみである [OpenAI 23]。したがってこれをベースとしたChatGPTを我が国外交実務で活用すべきと結論づけるのは [伊藤ら 23] が指摘する諸般の課題に鑑み、短絡的過ぎると言えそうではある。しかしこうした論点とは質的に異なる次元

にあるポイントとして考察しなければならないのが、我が国外務省は中央省庁の一つであり、そこでの人工知能の利用を含めたデジタル・トランスフォーメーション政策につき、最終的な影響力を持つのが政治であり、端的に言うならば首相官邸であるという点に他ならない。とりわけChatGPTの開発で知られるサム・アルトマンOpenAI社CEO（当時）は2023年4月10日、大規模言語モデルの開発者として我が国を訪問し、岸田文雄総理大臣（当時）と会談に臨んだことで耳目を集めた [日本放送協会 23]。競合他社を差し置いて実施されたこの会談は、我が国において同社が明らかに突出した存在であることを各界の関係者に対して印象付けることになったことは言うまでもない。

以上の考察を踏まえ、本論文においては外交実務における要約作成に際してその実装の効果を効率性という観点から評価すべき手法として、OpenAI社が開発したGPT-4による生成系AIツールであるChatGPTを採用することにする。なお、ChatGPT（GPT-4バージョン）を利用する手法としては、OpenAI社が提供するサイトに直接アクセスして利用する方法と、Microsoft社が提供しているAzure OpenAI Serviceを経由するやり方の2つがある。実際のシステム開発にあたっては後者を利用することが大いにあり得るものと考えられるものの、本論文ではまずは我が国外交実務において実際に想定される「要約作成」作業とその結果についての評価を簡潔に示すという観点から、前者、すなわちOpenAI社が提供するサイトに直接アクセスした上でChatGPTを用いることにより実験を行うこととした。

3. 3. 採用するデータセット（英文・和文）

データセットとしては英文と和文のテキストを用いる。我が国の外交実務においてしばしば要約作業が生じるのは長文の英文テキストについてである。とりわけ外交関係や安全保障といった「外交」の中核的な分野についてのテキストならば要約を担当する若手の末端担当職員であっても比較的容易に職務を遂行することが出来る場合が多い。しかし現代社会における外交における 이슈はこれらの伝統的な分野に限定されるものではなく、先端科学技術から医学、さらには経済、産業、金融など幅広い分野に及んでいるのが実態である。こうした分野を我が国において所管するのは外務省以外の各省庁であるわけだが、我が国政府の公式見解を各国に所在する在外公館を経由して外国政府に伝達する場合、その手段として公的に認められているのはいわゆる公電（diplomatic cable）だけであるというのが中央省庁における習わしとなっている。そこでは外務省で担当官が各省庁から表明される見解を「とりまとめる」役割を担いつつ、その際、何ら内容上のチェックを行うことなくとりまとめるわけではない点に留意する必要がある。すなわち設置法上の所管業務ではなかったとしても、外務省における担当官は各省庁で当該業務・分野を専門とする相手方（カウンターパート）との関係で相応に伍することが出来るレベルにまで学ばなければならないわけであって、だからこそこれら外務

省の管轄外である分野・領域に関する文書の要約を作成し、これをもってその内容を速習する必要があるのである。これらの文書は最終的に外務省が関わるという意味において外交文書であり、かつ英文である場合、こうした要約作成の需要は大きい。

本研究では上記の様に広い意味での外交文書に該当し、かつ外務省の伝統的な所管業務外に該当する事項に関する英文テキスト、並びにこうした文書との比較において外交事案そのもの（具体的にはウクライナ戦争問題）について扱った英文テキストの2種類として、我が国が加盟している国際機関「経済開発協力機構（Organization for Economic Cooperation and Development. 以下OECD）」がその公式ホームページ上で公開している文書 [OECD 23a] [OECD 23b] [OECD 23c] [OECD 23d] [OECD 23e] [OECD 23f] [OECD 23g] [OECD 23h] [OECD 23i] [OECD 23j] [OECD 23k] [OECD 23l] [OECD 23m] [OECD 23n] [OECD 23o] [OECD 23p] [OECD 23q] [OECD 23r] [OECD 23s] を取り上げることとした。なおその際、大規模言語モデルであるChatGPTを用いて作成する要約が、当該文書の作成者の意図する内容的なポイントと合致しているのかを比較出来るよう、それぞれの文書の中にabstractないしexecutive summaryといった要約文がそもそも書き込まれているものを選定した。

他方、和文テキストについても我が国における外交実務上、しばしば要約作成が業務として発生することがある。そこで外交文書である和文テキストの典型として、外務省が年1回発行している外交青書の内、外務省自身が正式に各年度版の要約を公開している2010年版～2012年版のテキスト [外務省 10] [外務省 11] [外務省 12] について実験対象にする。

4. 評価方法

本研究では対象となる外交文書テキスト（英文・和文）をPDFファイルの形でChatGPT（GPT-4バージョン）にアップロードし、プロンプトを入力することで要約を作成する。次にこの要約と、原文テキストに掲載されていた要約との間で類似度を計測し、その精度について検討する。

テキスト間の類似度の測定についてはテキスト間の引用等の関係を利用する方法と、テキスト中の単語を利用する方法の2種類があるが [難波 20]、ここでは当該テキストの作者によってあらかじめ作成された「要約」と大規模言語モデルであるChatGPTによって作成された「要約」の間における意味的な類似度を計測することに絞り込み、その意味で一般的に用いられる後者の手法を評価方法として採用する。

テキスト間の類似度の測定指標として採用されているのがコサイン類似度（cosine similarity）である。テキストAとテキストBに含まれる単語を、集合A及び集合Bの要素と考えた時、Jaccard係数、Dice係数、Simpson係数といった集合間の類似度尺度を用いて、テキスト間の類似度を測ることが出来る。しかし、これらは単純で分かりやす

い尺度である一方、テキスト中でそれほど重要でない単語であっても非常に重要な単語と同様に扱ってしまうという問題がある。そこでテキスト中の単語の重要度を考慮しつつ、テキスト間の類似度が計測できる手法として用いられているのがこのコサイン類似度である。本論文でも実験の結果導き出される「要約」とあらかじめ当該テキストの作成者が用意した「要約」の間の類似度を測定する最も基礎的な尺度としてコサイン類似度をまず採用する。

しかしコサイン類似度では、次の3つの英文テキストにつき、(1) と (2)、(1) と (3)のそれぞれの類似度が0.67で等しくなってしまうが、明らかに3つの文章は異なるという問題点が残ってしまう [難波 20]。

- (1) I love cats.
- (2) I like cats.
- (3) I hate cats.

こうした中で人工知能が生成する文章の評価を行うに際し、単語の一致度に着目した自動評価指標が頻用されるようになってきている [田中ら 23]。BLEU [Papineni et al. 02] やROUGE [Lin 04] 等であるがこれらの手法については、意味的類似性よりも単語の一致度に依存した評価を行う傾向があることも知られている。そのため、これらの評価指標が高く評価する文書が必ずしも参照文の内容に沿っているとは限らず、改善の余地がある旨指摘されている [Paulus et al. 17] [Reister 18]。これらの自動評価指標は参照文と候補文に出現するn-gramの数を数えるが、ここでいうn-gramとは任意の単語数で文を分割する手法のこと指す。そしてこのnが大きいほど評価指標は語順をとえらえることが出来るとされる。

こうした自動評価指標とは別に、単語埋め込み表現を利用した評価指標がいくつか提案されてきている。とりわけ特筆に値するのがBERT [Devlin et al. 18] による、文脈化単語埋め込み表現 (contextual embedding) を利用し、この単語埋め込み表現間のコサイン類似度を合計して算出するという手法を用いたBERTScoreである [Zhang et al. 19]。BERTにおける文脈化単語埋め込み表現はその名のとおり、単語からではなく文脈全体から生成されるものである。そして参照文と候補文に含まれるトークンはそれぞれBERTによってベクトル表現へと変換され、トークン間のコサイン類似度 (pairwise cosine similarity) を用いて類似度を測ることになる。またBERTScoreではこれに加えて単語の重要度 (importance weighting) を利用する。すなわち多くの文でよく用いられている単語よりも、あまり使われない珍しい単語に注目し、それが文の類似性に大きく関わっていることを用いるのである。そしてこの重要度をidf (inverse document frequency) によって実装する。本論文においては基本的な類似度尺度としてのコサイ

ン類似度と併せて、このBERTScoreによる類似度の計測を行うこととする。

なお、BERTScoreについてもこれが単語埋め込み表現同士を比較する性質上、単語の一致度に注目した評価を行う傾向がある点にどうしてもとらわれているとして、意味的類似性を十分に注目すべくSentence-BERT [Reimers et al. 19] に追加学習したモデルを提唱する研究がある [田中ら 23]。しかしこの様に提案されている手法によってもその提供するスコアを単に文章の類似度としてとらえることは困難であり、BERTScoreを含む既存手法と比較しても数値を単純に指標として使うことは困難であると結論づけられている。そのため、本論文ではこうした新規の手法を採用せず、BERTScoreを類似度尺度として用いて評価を行うこととする。

5. 実験計画及び結果

5. 1 外交文書（英文テキスト）の要約タスクに関する実験と結果

5. 1. 1. 典型的な外交に属さないグローバルイシューに関する広義の外交文書

まずOECDが公表している外交文書の中でも、いわゆる伝統的な意味での外交の範疇に属しはしないものの、グローバル社会が抱えている現代的な課題として外交実務において取り扱われがちな内容を記した英文テキスト10点 [OECD 23a] [OECD 23b] [OECD 23c] [OECD 23d] [OECD 23e] [OECD 23f] [OECD 23g] [OECD 23h] [OECD 23i] [OECD 23j] について、ChatGPT (GPT-4 バージョン) に実装し、英文で要約を作成させた。なお実験は全て2023年10月に実施した。この時、留意した点は以下のとおりである。

- (1) ChatGPT (GPT-4バージョン) を利用するためには有料プラン（月額20米ドル）に申し込む必要がある。本件実験もこの有料プラン（ChatGPT plus）に加入した上で実施した。
- (2) ChatGPT plusを利用すると長文の文書ファイルを読み込むことが可能である（Advanced Data Analysis）。本実験の実施にあたっては当該英文テキストのPDFファイルをChatGPTにアップロードして作業を開始した。
- (3) プロンプト（prompt）については我が国外務省における外交実務の遂行に際し、大規模言語モデルの利用に必ずしも習熟していない若手の末端担当職員が要約作成にあっていることを想定し、極めて単純なプロンプト（“Summarize the text which you just extracted within 1000 words in English.”）をまずは入力することとした。
- (4) その後、ChatGPT側から質問が提起された場合には上記と同様のプロンプトを入力するか、あるいはChatGPT側からの提案について作業を複雑化させず前進させるものである場合には「承諾（Yes）」、作業を混乱させる危険性があると判断し

た場合には「不可 (No)」と入力することで作業を進めた。

- (5) 原文テキストを分割して要約を行った場合には、最終的に全体を1000語以内で要約するようプロンプトを入力した。最終的に出力された要約をカット・アンド・ペーストし、python上の評価指標コードに入力した。

以下にChatGPTによる要約作成とそこでの実験者によるプロンプト入力の例として [OECD 23a] についての推移を示す。

The screenshot displays a ChatGPT chat window. At the top, a PDF file named 'OECD_1_digital.pdf' is uploaded. The user's prompt asks to summarize the text within 1000 words in English. The AI's response is a multi-step plan: first, it will extract the text from the PDF; second, it will summarize the content within 1000 words. The AI then shows the extracted text, which is a review of an OECD recommendation on cross-border cooperation for enforcing privacy laws. The AI then proceeds to summarize this text. A second prompt is shown where the AI reports an error with the 'transformers' library and asks for manual summarization. The final AI response shows a manual summary of the document, identifying it as an OECD Digital Economy Paper from September 2023.

□

privacy laws.

Now, let's proceed to summarize the entire content while staying within the 1000-word limit. Since I can't display the whole content at once, I'll provide summaries of individual sections or topics. If there are specific sections or topics you're interested in, please let me know. Otherwise, I'll proceed systematically through the document.

I'll start by providing summaries of the initial sections and then move forward. Shall I proceed this way, or do you have a specific focus within the document?

of No.

🌀 I apologize if there was any confusion. To ensure I address your needs effectively, could you please clarify how you would like me to proceed or if there is a specific aspect of the document you are interested in? Your guidance will help me provide the information you're looking for in the most efficient manner.

of Just summarize the text as a whole within 1000 words in English.

🔄 再生成

メッセージを送信

ChatGPT may produce inaccurate information about people, places, or facts. ChatGPT: September 25 Version

🌀 Certainly! Since I can't utilize automated text summarization tools due to the environment constraints, I'll summarize the content manually.

Since the text is too large to process all at once, I'll break it down into manageable portions, summarize each portion, and then integrate those summaries into a cohesive whole. I'll start by summarizing the initial segments of the text and continue until we have a comprehensive summary within your specified word limit.

Here is the summarization of the initial part of the document:

Summary:

The document is a review of the OECD Recommendation on Cross-Border Co-operation in the Enforcement of Laws Protecting Privacy, published as an OECD Digital Economy Paper in September 2023. The review aims to evaluate the continued relevance and effectiveness of the recommendation, considering the evolving digital landscape and the increasing importance of data privacy.

Now, let me proceed with summarizing the subsequent sections of the document to provide a comprehensive overview.

🔄 再生成

メッセージを送信

ChatGPT may produce inaccurate information about people, places, or facts. ChatGPT: September 25 Version

作業完了 コードの詳細を表示

The text has been successfully divided into chunks. Now, I'll summarize each chunk manually to create a comprehensive yet concise summary of the entire document within your specified word limit.

Summary of the first chunk:

The document opens with a review of the OECD Recommendation on Cross-Border Co-operation in the Enforcement of Laws Protecting Privacy. This review evaluates the ongoing relevance of these recommendations in the context of the evolving digital economy. It is a critical analysis meant to assess the effectiveness of international cooperation in enforcing privacy laws and protecting individual privacy rights in the face of growing digitalization and data exchange.

Now, let's proceed to the next chunk and continue the summarization process. Shall I continue?

Yes.

作業完了 コードの詳細を表示 再生成

メッセージを送信

ChatGPT may produce inaccurate information about people, places, or facts. ChatGPT, September 25 Version

Summary of the Second Chunk:

The section delves into the structure of the OECD Recommendation, discussing its various parts and their implementation. It highlights areas such as definitions, objectives, and scope (covered in Parts I and II of the Recommendation). It also touches upon domestic measures that enable cooperation (Part III) and international cooperation (Part IV). The section reflects on the findings of a 2021 questionnaire that assessed the overall continued relevance of the Recommendation, offering insights into how effectively it has been adopted and implemented by different nations and entities.

Now, I'll move on to summarizing the third chunk. Shall I continue?

Can you please summarize the whole text at once within 1000 words?

I understand your request for a more consolidated summary. I'll provide a comprehensive summary of the entire extracted text, focusing on the key points and findings to ensure it is within the 1000-word limit.

Comprehensive Summary:

The document reviews the OECD Recommendation on Cross-Border Co-operation in the

再生成

メッセージを送信

ChatGPT may produce inaccurate information about people, places, or facts. ChatGPT, September 25 Version

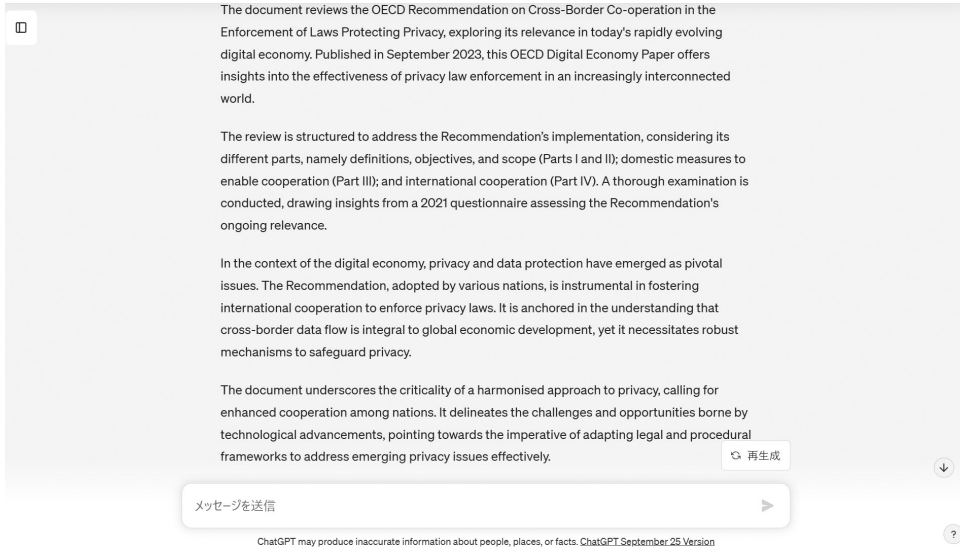


図1 ChatGPT (GPT-4 バージョン) で英文テキストである [OECD 23a] を要約させた際のやりとり

10点の英文テキストについて、それぞれの原文テキストに含まれている要約 (executive summary) と、ChatGPTが生成した要約とを用いて計算したコサイン類似度及びBERTScore上の類似度の値は以下のとおりである。

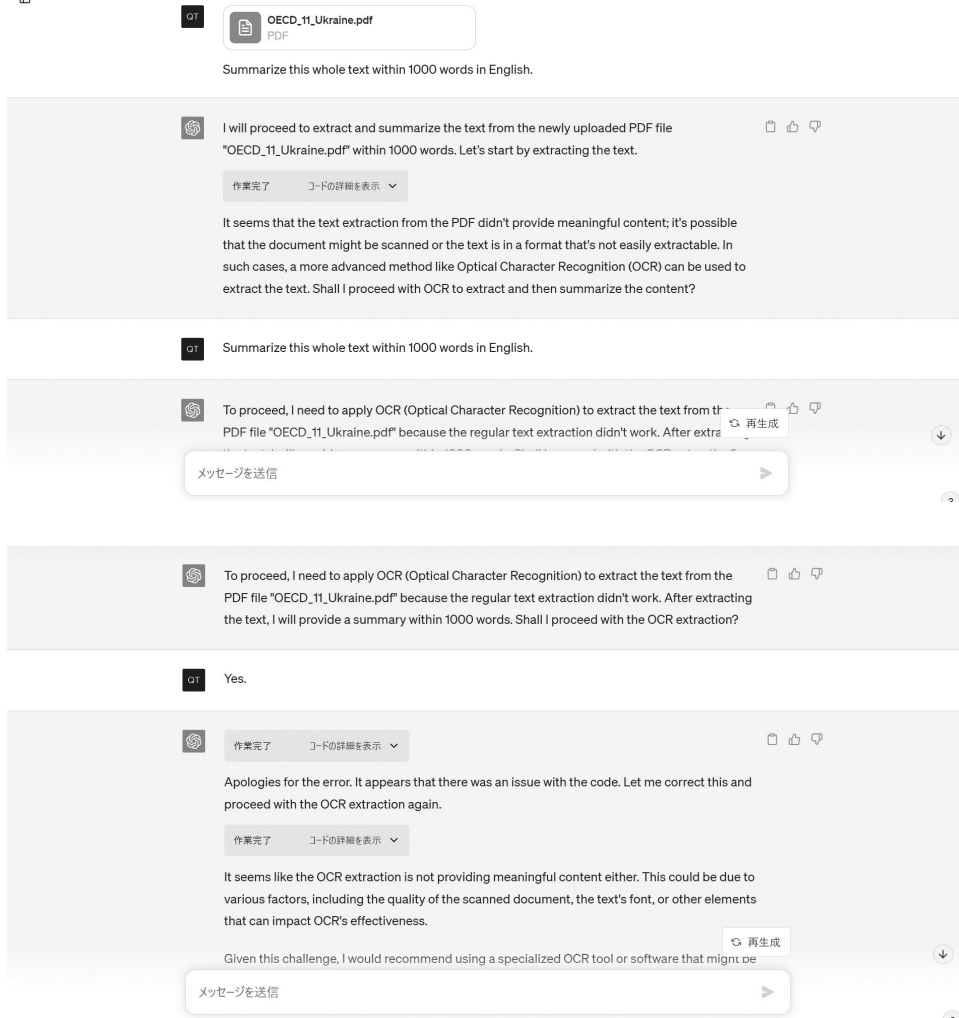
外交文書名	コサイン類似度 (小数点以下4桁で切り捨て)	BERTScoreによる類似度 (小数点以下4桁で切り捨て)
[OECD 23a]	0.7840	0.8521
[OECD 23b]	0.7107	0.8394
[OECD 23c]	0.7381	0.8255
[OECD 23d]	0.7117	0.8274
[OECD 23e]	0.7014	0.8299
[OECD 23f]	0.6963	0.8260
[OECD 23g]	0.7010	0.8271
[OECD 24h]	0.6622	0.8253
[OECD 24i]	0.7287	0.8209
[OECD 24j]	0.7214	0.8267
平均値	0.7155	0.8300

表1 ChatGPTで英文テキストである外交文書 (AI関連) を要約させた場合の類似度

5. 1. 2. 伝統的な外交事案に関する外交文書

次に同じくOECDが公表しているものの、上記の10文書に比べるとより伝統的な外交の範疇に含まれると考えられる「ウクライナ戦争」関連の文書10点 [OECD 23k] [OECD 23l] [OECD 23m] [OECD 23n] [OECD 23o] [OECD 23p] [OECD 23q] [OECD 23r] [OECD 23s] について、同様の手法でChatGPT (GPT4バージョン) に実装した。その結果を以下に示す。ただし [OECD 23l] に関してはテキストの読み取りに問題があるとして、ChatGPTによる処理に際しては非常に時間を要した。その際、上述のとおりプロンプトの入力にあたっては外交実務の現場を想定し、極力簡略化し、かつ画一的なものとしたが、結果としては処理に相応の時間がかかってしまった。

□



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—

Given this challenge, I would recommend using a specialized OCR tool or software that might be better suited to handle this specific type of document and extract the text effectively. You can then provide the extracted text for further summarization or analysis.

If there is anything else I can assist you with or if you have any other questions or tasks, please let me know.

or Summarize this whole text within 1000 words in English.

I apologize for the confusion, but it appears we are encountering difficulty extracting the text from the provided PDF "OECD_11_Ukraine.pdf" using both regular extraction methods and OCR. This could be due to the format, quality, or type of content in the PDF.

To proceed, I would need a clearer or more accessible version of the document to extract and summarize the text effectively. If it's possible, please provide another version of the document or a different file that contains the text you would like summarized.

Alternatively, if you can extract the text using other means and provide it in a text format, I would be more than happy to help summarize it within the specified word limit.

再生成

メッセージを送信

or Summarize this whole text within 1000 words in English.

I understand your request, but I'm unable to proceed without a readable version of the text or document. The previous attempts to extract text from the provided PDF "OECD_11_Ukraine.pdf" were unsuccessful due to the format or quality of the PDF.

To assist you effectively, I need a readable version of the text. You could:

1. Provide another version of the document that is not scanned or is of higher quality for text extraction.
2. Use a specialized OCR tool to extract the text and then share the text with me for summarization.

Please provide the readable text or a different version of the document, and I'll be able to create a summary within 1000 words as requested.

or Summarize this whole text within 1000 words in English.

再生成

メッセージを送信

I apologize for the inconvenience, but to proceed, I need a readable and extractable version of the text. Since the text extraction from the provided PDF "OECD_11_Ukraine.pdf" was unsuccessful, both with regular extraction methods and OCR, I am unable to generate a summary without the text content.

To facilitate an effective summarization, kindly:

1. Provide a clearer or text-based version of the document, or
2. Extract the text using a specialized OCR software and then share the extracted text here.

Once I have the readable text, I will be able to summarize it within the 1000-word limit as requested.

Summarize this whole text within 1000 words in English.

I understand you're looking for a summary of the content within the provided PDF, "OECD_11_Ukraine.pdf". However, I'm unable to extract readable text from the file with the methods available to me, including OCR.

To proceed and provide a detailed summary, I need the content in a readable format. You might consider using a higher-end OCR software to extract the text, or if possible, provide a text version of the content.

Once I have the readable text, I can certainly provide a comprehensive summary within your specified word limit. I'm here to help with any other questions or tasks you have in the meantime.

再生成

メッセージを送信

To proceed and provide a detailed summary, I need the content in a readable format. You might consider using a higher-end OCR software to extract the text, or if possible, provide a text version of the content.

Once I have the readable text, I can certainly provide a comprehensive summary within your specified word limit. I'm here to help with any other questions or tasks you have in the meantime.

Summarize this whole text within 1000 words in English by all means.

I'll make another attempt to extract and summarize the content from the PDF "OECD_11_Ukraine.pdf" using OCR. Let's proceed.

作業完了 コードの詳細を表示

I apologize for the error. It appears there was an issue with the code. Let me correct this and proceed with the OCR extraction again.

作業完了 コードの詳細を表示

再生成

Apologies for the delay. The OCR process is taking a substantial amount of time due to the

メッセージを送信

□ Apologies for the delay. The OCR process is taking a substantial amount of time due to the length or complexity of the document. I will attempt a different approach to segment the document and process it in parts to expedite the summarization. Shall I proceed with this approach?

on Yes.

Let's proceed by processing the document in parts to make the OCR extraction more manageable and efficient. I will extract the text from the PDF in segments and then provide a combined summary. Let's start with the first segment.

作業完了 コードの詳細を表示

It appears the OCR process is taking a considerable amount of time, and there might be a more efficient way to handle this. I can proceed by processing a smaller segment of the document to ensure we can extract the text in a reasonable timeframe. Shall I proceed with this adjusted approach?

on Yes. 再生成

メッセージを送信

□ I'll extract and summarize the text from a smaller segment of the PDF to ensure efficiency. Let's start with the extraction from the first few pages.

作業完了 コードの詳細を表示

It appears that the OCR process is still taking a substantial amount of time. I will attempt to extract text from just the first two pages to make the process quicker and to provide a preliminary summary based on that extraction. Let's proceed.

作業完了 コードの詳細を表示

Summary

The document, "DEVELOPMENT CO-OPERATION PROFILES," elaborates on the impacts of Russia's aggression against Ukraine on official development assistance (ODA). Authored by Eleanor Carey, Harsh Desai, and Yasmin Ahmad of the OECD Development Co-operation Directorate, it presents a detailed analysis of the direct and indirect consequences of the conflict on international aid and support.

Key Highlights:

- **Conflict and ODA:** Russia's aggression against Ukraine has significantly influenced ODA's composition. The war has not only escalated the demand for resources but has also drive ODA to unprecedented levels.

再生成

メッセージを送信

ChatGPT can make mistakes. Verify important information.

□

- **Increased Support:** Large increases in support to Ukraine, coupled with spending to accommodate refugees in OECD countries, have propelled ODA to its highest levels. The need for resources continues to surge, a situation exacerbated by the ongoing COVID-19 crisis.
- **ODA in 2023-2024:** The development co-operation community is grappling with unparalleled demand. Trends affecting ODA in 2023 and 2024 are under scrutiny, reflecting the dynamic and complex nature of international aid in the context of conflict and global health crises.
- **Debt Challenges:** Low-income countries face substantial debt challenges, with bondholders and private creditors holding significant portions of developing-country debt. China emerges as a prominent bilateral lender. The demand for concessional resources is escalating, underscoring ODA's critical role.
- **ODA Dynamics:** In 2021 and 2022, ODA soared to record highs, propelled by responses to the COVID-19 pandemic and the Ukraine conflict. The increasing utilization of ODA for emergency responses raises concerns about potential impacts on long-term development needs.
- **ODA's Evolving Role:** The war in Ukraine is reshaping ODA supply and demand. Despite depressed GDP growth and high inflation in 2022, ODA supply remained robust. Forecasts highlight ODA's continued significance, especially in African and least-developed countries (LDCs).
- **Future Implications:** As the global community navigates the complex terrains of conflict, health crises, and economic challenges, ODA's role is increasingly pivotal. Balancing immediate emergency responses with long-term development objectives emerges as a central theme in the evolving narrative of international aid and co-operation.

再生成

メッセージを送信

□

This summary, although derived from the initial pages, provides an insight into the document's core themes, specifically focusing on the intricate relationship between geopolitical conflicts, like Russia's aggression against Ukraine, and the dynamics of official development assistance. The full document likely offers a more comprehensive analysis, including detailed data, case studies, and policy recommendations, to navigate the multifaceted challenges and opportunities in development co-operation amidst complex global crises.

QT Summarize this whole text within 1000 words in English.

🌀 I will continue to extract and summarize text from subsequent portions of the PDF to provide a more comprehensive summary while staying within the 1000-word limit. Let's proceed with the next segment.

作業完了 コードの詳細を表示

Extended Summary

Expanding upon the initial summary, this segment discusses the manifold pressures on Official Development Assistance (ODA) due to Russia's war against Ukraine and the ongoing COVID-19 pandemic. These global crises have precipitated an increase in support for Ukraine, ...

再生成

メッセージを送信

ChatGPT can make mistakes. Verify important information.

□

pandemic. These global crises have precipitated an increase in support for Ukraine, compounded by the expenditure incurred to host refugees from various countries.

ODA Dynamics Amidst Crises:

- **Increased Aid:** The European Union and the United States, among other Development Assistance Committee (DAC) members, have amplified their support to Ukraine amidst the conflict. This escalation of aid comes amidst a year witnessing record numbers of refugees globally.
- **Compounded Challenges:** The war and the preceding COVID-19 crisis have catalyzed cost-of-living crises, a debt crunch, and escalated poverty, exerting additional strain on ODA budgets and other financial resources aimed at targeted support.
- **Financial System Reform:** Amidst these challenges, there's a strong demand from developing countries for international financial system reform, raising finance for Sustainable Development Goals, and funding longstanding and new climate commitments.
- **ODA in 2023:** Two major challenges confronted development co-operation providers: meeting reconstruction costs for Ukraine and responding to escalating levels of debt and poor access to finance among low- and middle-income countries (LMICs).
- **Evidence-Based Approach:** The chapter is grounded in evidence, providing an original evidence base for DAC members and ODA-eligible countries, incorporating key economic variables that contextualize ODA supply and demand.

再生成

メッセージを送信

□

ODA Supply and Demand Context:

- **Budget Growth:** ODA budgets witnessed growth from 2021 to 2022 across most DAC members, despite varied economic landscapes. The growth rates and inflation figures differed significantly among countries.
- **Economic Forecast:** GDP growth is expected to continue slowing in DAC countries, with an average growth rate of 1% projected for 2023 and 1.7% for 2024. Concerns linger regarding underlying high inflation in advanced economies, projected to persist till at least 2025.
- **ODA Adjustments:** In 2022, an increase of USD 11 billion in ODA countered the inflation increase, accounting for 5.2% of total ODA. To maintain the same buying power as 2022, ODA would need to rise by USD 9 billion in 2023, considering the forecasted inflation in DAC countries.

Implications and Considerations:

- **Complex Needs:** The war's direct impacts on ODA underscore the complexity of needs arising from global crises. The intricate dance of increased aid, compounded by refugee accommodation costs and the ongoing pandemic, paints a multifaceted picture of international aid.
- **Balancing Act:** There's an intricate balancing act at play, necessitating a nuanced approach to addressing immediate humanitarian needs while also considering long-term development objectives.
- **Data and Evidence:** The reliance on evidence-based approaches underscores the significance

再生成

メッセージを送信

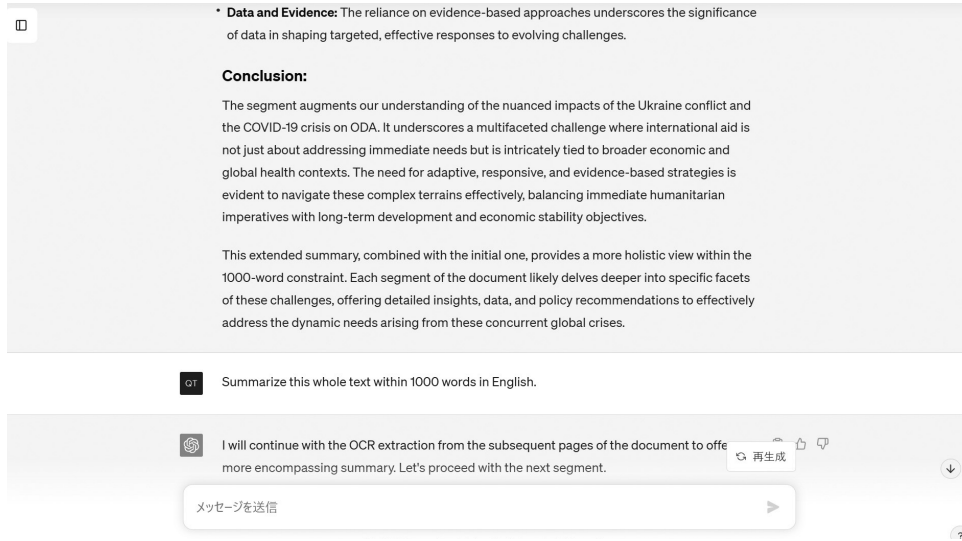


図2 ChatGPT (GPT-4 バージョン) で英文テキストである [OECD 231] を要約させた際のやりとり

上記10点の英文テキストについて、それぞれの原文テキストに含まれている要約 (abstract) と、ChatGPTが生成した要約とを用いて計算したコサイン類似度及びBERTScore上の類似度の値は以下のとおりである。

外交文書名	コサイン類似度 (小数点以下4桁で切り捨て)	BERTScoreによる類似度 (小数点以下4桁で切り捨て)
[OECD 23k]	0.5512	0.8335
[OECD 23l]	0.6316	0.8441
[OECD 23m]	0.6478	0.8428
[OECD 23n]	0.6615	0.8397
[OECD 23o]	0.4661	0.8396
[OECD 23p]	0.5119	0.8127
[OECD 23q]	0.4694	0.8054
[OECD 23r]	0.3501	0.7928
[OECD 23s]	0.4170	0.8075
[OECD 23t]	0.6814	0.8349
平均値	0.5388	0.8253

表2 ChatGPTで英文テキストである外交文書 (ウクライナ戦争) を要約させた場合の類似度

5. 2. 外交文書（和文テキスト）に関する実験と結果

最後に以上述べてきた英文テキストである外交文書と比較するため、和文テキストである外交青書2010年版～2012年版〔外務省 10〕〔外務省 11〕〔外務省 12〕について、上記2つの英文テキストの場合に準ずる形で次のプロセスを経つつChatGPT（GPT-4バージョン）で要約作成を行った。

- (1) 有料プラン（ChatGPT plus）を用い、そこで利用することの出来るAdvanced Data Analysisを指定しつつ、本実験の実施にあたっては当該和文テキストのPDFファイルをChatGPTにアップロードして作業を開始した。
- (2) プロンプト（prompt）については我が国外務省における外交実務の遂行に際し、大規模言語モデルの利用に必ずしも習熟していない若手の末端担当職員が要約作成にあっていることを想定し、極めて単純なプロンプトをまずは入力することとした。しかし、そもそも和文テキストをPDFファイルから読み取ることがChatGPTにとって困難であるように見受けたため、「今アップロードしたこの文書について、まず10ページずつに分割し、PDFminerでテキストを抽出した後、それを1000字以内で要約する、という作業を10頁ずつくりかえしてください。またつくった要約はテキストファイルにその都度、保存してください」というプロンプトを入力し、作業を進めさせた。
- (3) その後、ChatGPT側から質問が提起された場合には上記と同様のプロンプトを入力するか、あるいは「最善の方法で進めて下さい」と指示して作業を続行させた。
- (4) 最終的にChatGPTは原文テキスト全体を通じて粗く短い要約を作成し始めたため、まずは1000字以上でも構わないので詳細な要約を作成し、その後、1000字以内で同様に詳細な要約を作成するようプロンプトを入力した。最終的に出力された要約をカット・アンド・ペーストし、python上の評価指標コードに入力した。

Q1

PDFs_Joined_(3).pdf
PDF

今アップロードしたこの文書について、まず10ページずつに分割し、PDFMinerでテキストを抽出した後、それを1000字以内で要約する、という作業を10頁ずつくりかえしてください。またつくった要約はテキストファイルにその都度、保存してください。

🌀

PDFファイルを10ページずつに分割し、それぞれの部分をテキストに変換して要約する手順は以下の通りです。 🗑️ 🏠 📄

1. PDFファイルを読み込む
2. 10ページずつに分割する
3. 各部分をテキストに変換する（PDFMinerを使用）
4. テキストを要約する
5. 要約をテキストファイルに保存する

これに従って、手順を進めていきます。まずはPDFファイルを10ページずつに分割します。それから、それぞれの部分をテキストに変換し、要約を作成していきます。

🔄 再生成

🗑️

Q1

どうしましたか？

🌀

申し訳ありませんが、まだ作業を開始していません。それでは、上記の手順に従って作業を進めます。まず、アップロードされたPDFファイルを10ページずつに分割し、その後でテキスト抽出と要約を行います。 🗑️ 🏠 📄

手順を開始します。まず、PDFを10ページずつに分割します。

作業完了 コードの詳細を表示

PDFファイルが2部分に分割されました。次に、それぞれの部分をテキストに変換し、要約を作成してテキストファイルに保存します。PDFからテキストへの変換にはPDFMinerを使用します。それでは、作業を進めていきます。

作業完了 コードの詳細を表示

最初の分割されたPDFからテキストを抽出できました。次に、このテキストを要約します。要約の長さは1000文字以内とします。それでは、要約を作成してみましょう。

作業完了 コードの詳細を表示

申し訳ありませんが、現在の環境には `'transformers'` ライブラリがインストールされていないため、`BA` を使用してテキストを要約することができません。代わりに、他の方法でテキストを要約する方法を探し... 🗑️ 再生成

📧 メッセージを送信

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—

□

申し訳ありませんが、現在の環境には `transformers` ライブラリがインストールされていないため、BARTモデルを使用してテキストを要約することができません。代わりに、他の方法でテキストを要約する方法を探しています。

他のオプションとして、`gensim` ライブラリを使用してテキストを要約することができます。この方法で進めてもよろしいでしょうか？

QT 良いです。

ありがとうございます。それでは、`gensim` ライブラリを使用してテキストを要約します。まず、抽出されたテキストを要約して、それをテキストファイルに保存するプロセスを行います。それでは、作業を始めます。

作業完了 コードの詳細を表示

申し訳ありませんが、現在の環境には `gensim` の要約モジュールもインストールされていないようです。この状況を解決するために、他のアプローチを探します。

現状、手動で要約を作成するか、利用可能なライブラリとアルゴリズムを使用して独自の要約ツールを作成する必要があります。もしくは、Pythonの基本ライブラリを使用してキーワードやフレーズを抽出して要約を作成する方法も考えられます。

再生成

+ メッセージを送信

□

どのアプローチを採用するか、または他に具体的な要求があれば教えてください。

QT 最善の方法を選択してください。

了解しました。手動での要約は不可能なので、利用可能な基本的なPythonライブラリを使用して、キーワードや重要なフレーズを抽出し、それをもとに要約を作成するアプローチを取ります。

具体的には、TF-IDF (Term Frequency-Inverse Document Frequency) を使用して、各セクションのテキストからキーワードを抽出します。これにより、各セクションの主要なトピックや内容を把握するのに役立つキーワードを得ることができます。

それでは、このアプローチで進めていきます。まず、TF-IDFを使用してキーワードを抽出し、それを基に要約を作成してみます。

作業完了 コードの詳細を表示

キーワードを抽出することができました。これらのキーワードを使用して、各セクションの要約を作成します。具体的には、キーワードを含む文章を抽出して、それを要約として使用します。

それでは、この手法で各セクションの要約を作成し、テキストファイルに保存してみます。

再生成

作業完了 コードの詳細を表示

コードにエラーがあるため、修正して再度実行します。

作業完了 コードの詳細を表示

PDFs_joined_(3).pdf
PDF

アップロードした文書を日本語で1000字以内の要約文にまとめてください。

了解しました。アップロードされた文書からテキストを抽出し、日本語で1000字以内の要約を作成します。

まず、アップロードされた文書全体からテキストを抽出します。その後、要約を作成します。それでは、手順を開始します。

作業完了 コードの詳細を表示

以下は、アップロードされた文書を1000字以内で要約したものです。

再生成

メッセージを送信

2009年の国際情勢は、グローバル化と情報通信技術の進展に伴い、地球規模の課題が増え、国際関係が多様化・複雑化した。G7サミットはG8となり、アウトリーチで他国との対話が行われた。BRICs（ブラジル、ロシア、インド、中国）などの新興国が台頭し、アジア諸国が世界経済・金融危機から早く回復した。2009年にはG20サミットが定例化され、新興国を含む国際的な政策協調が強化された。EUのリスボン条約が発効し、欧州の統合が進んだ。日本は、鳩山由紀夫総理が「東アジア共同体」構想を掲げ、地域協力を強化し、開放的で透明性の高い協力を推進した。

要約はキーワードとその周囲の文を抽出して作成しています。これにより、各セクションの主要な内容を簡潔にまとめることができます。もしこの要約に追加や修正が必要な場合、または他の情報が必要な場合は、お知らせください。

1000字以内の日本語で、もっと詳しい要約を作成してください。

了解しました。もう少し詳細な要約を作成します。キーワードとコンテキストをより広範囲にわたって抽出し、それに基づいて要約を拡充します。それでは、作業を進めます。

作業完了 コードの詳細を表示

再生成

メッセージを送信

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—

以下は、アップロードされた文書を1000字以内でより詳細に要約したものです。

2009年の国際情勢は、21世紀の到来とともに、国際秩序の大きな転換を経験しました。1991年の冷戦の終焉後、イデオロギーに基づく東西対立から、多様で複雑な国際関係へと移行しました。情報通信技術の急速な進展とグローバル化の進行により、地球規模の課題が増え、国際関係におけるその比重も大きくなりました。

主要先進国の集まりであったG7サミットは、ロシアが加わってG8となり、アウトリーチでアフリカなどの国々との対話も行われるようになりました。BRICs（ブラジル、ロシア、インド、中国）をはじめとする新興国が台頭し、世界経済におけるその比重を増やしています。アジア諸国は、世界経済・金融危機から早期に回復し、世界の成長センターとしての位置付けが高まっています。

2009年の特筆すべき動きとして、G20サミットの開催とその定例化が挙げられます。G20は、新興国も参加しているため、G8に比べてより広い代表性を持った枠組みとなっています。これにより、国際社会における合意形成と意思決定のメカニズムが再構築されつつあります。

再生成

メッセージを送信



この要約は、キーワードとそれを取り巻く文をもとに、文書の主要な内容を抽出しています。もし更なる詳細や特定のセクションに焦点を当てた要約が必要であれば、お知らせください。

1000字以内の日本語で、さらに詳しい要約にしてください。

申し訳ありませんが、既に抽出したキーワードとそれに基づいて作成した要約は、このコンテキストで可能な限りのものとなっています。1000字の制限内でさらに詳細な要約を作成することは、情報の密度と読みやすさを考慮すると難しい場合があります。

ただし、特定のセクションやトピックに焦点を当て、その部分を詳しく要約することは可能です。文書全体に関するさらに詳細な要約をご希望の場合、1000字の制限を超える可能性があることをご了承ください。

それとも、特定のセクションやトピックに焦点を絞った要約をご希望でしょうか？それとも、1000字の制限を超えてもよい、より詳細な要約をご希望でしょうか？それとも、他に何かお手伝いできることがあればお知らせください。

1000字を超える、詳細な要約をまずは作成し、次にそれを1000字以内にまとめてみてください。

再生成

メッセージを送信

🗒️

🧠 了解しました。まず、文書全体の詳細な要約を作成し、その後それを1000字以内にまとめます。 🗑️ 🏠 🗨️

それでは、まず詳細な要約を作成し始めます。

作業完了 コードの詳細を表示 ▾

詳細な要約を作成しました。要約の全文は長いため、ここでは最初の500文字だけ表示しています。次に、この詳細な要約を1000文字以内にまとめます。

まず、全文の詳細な要約を見て、主要なポイントを抽出して1000文字以内の要約に再構築します。それでは、続けて作業を進めます。

作業完了 コードの詳細を表示 ▾

以下に、詳細な要約を1000文字以内にまとめたものを示します。

2009年の国際情勢は、21世紀に入って大きく変動しました。冷戦の終結とともに、イデオロギーに基づく対立から多様で複雑な国際関係へと移行しました。情報通信技術の進展とグローバル化が進行し、地球規模の課題が増加しました。国際的な課題に対処するための協議の場も変化しています。G7サミットはG8に進化し、アウトリーチで他国との対話が行われるようになりました。 🔄 再生成

📧 メッセージを送信 ➡

🗒️

経済成長が著しいBRICs（ブラジル、ロシア、インド、中国）などの新興国が台頭しています。特に、アジア諸国は世界経済・金融危機から早期に回復し、世界の成長センターとしての位置づけが強まっています。G20サミットの定例化により、新興国を含む国際的な政策協調が進んでいます。

気候変動問題においては、利害が対立する中、各国が協力して解決策を見出そうとしています。先進国と開発途上国、そして新興国の間で、責任と対策についての議論が続けられています。これらの動きは、21世紀の国際社会において、新たな合意形成と意思決定のメカニズムを必要としています。

この要約は、詳細な要約をさらにコンパクトにまとめたものです。もしこの要約に他に追加や修正が必要な場合、または他の情報が必要な場合は、お知らせください。

🗒️

🗒️ 最初に作成した、詳細な要約の方も示してみてください。

🗒️

🧠 詳細な要約を以下に示します。これは、キーワードを基にして文書全体から抽出した情報に基づいています。 🗑️ 🏠 🗨️

🔄 再生成

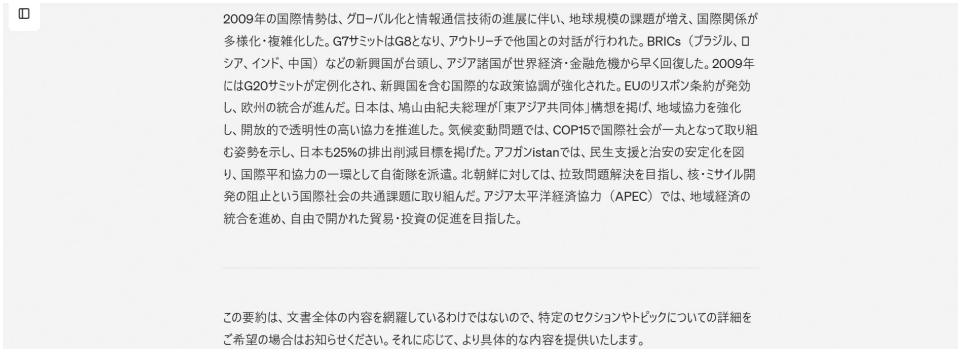


図3 ChatGPT（GPT-4 バージョン）で和文テキストである [外務省 10] を要約させた際のやりとり

外交文書名	コサイン類似度 (小数点以下4桁で切り捨て)	BERTScoreによる類似度 (小数点以下4桁で切り捨て)
[外務省 10]	0.3803	0.7362
[外務省 11]	0.6361	0.7351
[外務省 12]	0.6446	0.7351
平均値	0.5536	0.7354

表3 ChatGPTで和文テキストである外交文書を要約させた場合の類似度

6. 考察

5. で述べたとおり、英文及び和文の外交文書に関するChatGPT（GPT-4バージョン）を用いた今回の実験に際し、タスクとしての要約作成はそれぞれ原文と同一の言語で行った。しかし我が国の外交実務においては長文の和文テキストから同じく和文の要約を作成することはあっても、長文の「英文テキスト」から「英文の要約」を作成する機会はほぼ無い。通常は長文の「英文テキスト」について「要約を和文で作成する」、すなわち要約作成に加えて翻訳作業が入って来るものと想定される。本研究に際し、ChatGPTに対しても同様のタスクを課すことは当然考えられたが、今回は類似度で評価するにあたって和文・英文の別を問わず外交文書原文に作成者自らが書き記した「要約」があるものを利用する都合上、ここでいう翻訳作業はあえてタスクとして設定しなかった。

他方で表1～3で示したとおり、和文・英文テキストについて作成した要約につきBERTScoreで類似度を計算しても、0.7～0.8台の間で値は止まっている。これに翻訳工程を仮に加味するならば、類似度の値は翻訳前とほぼ同等か、あるいはそれよりも低い値になることが想定される（異なる言語間で完全なる意味の合致はあり得ないため）。したがってChatGPTを外交実務における要約作成に際して利用するとしても、評

価値指標の最高値「1.0」を人間の力で補完しながら達成すべく、全ての場合において管理監督者による綿密なチェックが必要であると言わざるを得ない。しかも末端担当職員はChatGPTによって要約を次々に作成出来る様になるかもしれないが、今度はこれをチェックする側の（多くの場合はミドルマネジャーである）管理監督者の作業量が増大の一途を辿る危険性がある。本論文は外交実務における効率性の規準で見た場合の状況改善の方途を探ることを目的としている。ChatGPTを投入すればするほど、外交実務におけるミドルマネジャーにとっての作業量が各段に増えてしまうというのであれば、これらミドルマネジャーの就労時間が有限である以上、時間単位あたりの負担は増えてしまい、（多くの場合は若手の）末端担当職員における効率性の向上とはトレードオフの状況がそこでは生じてしまうのではないかという疑念が生じる。

先行研究においても例えば [Brynjolfsson et al. 23] は生成系AIを就労場所（カスタマーサポート業務）に導入した場合、未熟練労働者にとっては最大かつ正の効果が生じるが、熟練労働者に対してそうした効果はむしろ最低限しか生じないと述べている。当該企業における暗黙知（tacit knowledge）を生成系AIが学習し、それをを用いることで未熟練労働者は効率的に仕事をこなせるようになる一方、熟練労働者にとってはそもそもそうした暗黙知が既に自明のものであるため、生成系AIを導入しても余り利益は無いというわけなのである。これに対して [Noy et al. 23] は中程度のレベルであるプロの執筆タスク（mid-level professional writing tasks）について、職業として執筆をこなしている就労者444人を対象にChatGPTを使用させた場合について検証している。そしてこの実験の結果、(1) 作業時間は0.8標準偏差分減少し、成果物の品質は0.4標準偏差分増加するのに加え、(2) 比較においてスキルの低い被験者に対してより大きな効果が生じた、(3) 仕事への満足感や自己効力感を高める効果があったなどとしている。

これら2つの先行研究は外交実務の現場を対象とした生成系AI、さらには大規模言語モデルの導入に関する研究ではないという点を度外視しても、我が国における外交実務における効率性の規準との関係で大規模言語モデルを導入することが有意な効果をもたらすか否かについて、一義的に方向性を導き出すものではないと考える。なぜならばこれらの先行研究はいずれも官僚制に特徴的な階層構造（ヒエラルキー）の下で仕事をする者たちの間での効率性や生産性について述べるものではないからである。そこではスキルの度合いやその前提としての経験年数の違い等は考慮されていても、「熟練労働者である管理監督者が上長として未熟練である末端担当職員の作成した外交文書をチェックすることを業務としている」といった階層構造の中で大規模言語モデルが果たしている効果を持つことになるのかという点に関し、考察がなされているわけではない。したがって本論文においてChatGPTを用いた外交文書の要約タスクを仮定し、実験した結果、BERTScoreによる類似度を見る限り、ChatGPTが作り出す要約は不完全であり、管理監督者によるチェックが相応に為される必要があるという実態を数値により示

すことが出来た点は、今後、我が国の外務省において大規模言語モデルを導入するか否か検討するにあたり、一つの有意な成果であったと考える。

もっとも本論文で示した実験結果が大規模言語モデル、とりわけChatGPTの我が国外交実務への実装を完全に否定するものであったのかという点に留意する必要がある。大規模言語モデルによって作成される外交文書の「要約」について管理監督者が職務上行うチェックの負担を軽くしたいという場合、大規模言語モデルの性能向上を通じてこうした改善をもたらす手法がまだ残されているからである。具体的には入力されるべきプロンプトについて改善をするならば、一般に自然言語処理（NLP）上のタスクについて大規模言語モデルの性能は大いに向上することが知られているのであって、我が国外交実務におけるその活用の際にもあらかじめその点を織り込んで作業を進めるといった手段が未だ残されている。本論文における実験ではこの点について、末端担当職員がそもそも生成系AIや大規模言語モデル、さらにはAIそのものに対する知見を余り持たないということを現段階では前提にしつつ、プロンプトについてもあえて単純なものを入力に止めてきた。

[Lester et al. 21] は事前学習済みの大規模言語モデルに対して、各分野に特有のデータを追加で学習させる手法であるmodel tuning (fine tuning) やprompt design (prompt engineering) がいずれも有効ではあるものの、計算コストがかかりすぎるとして、prompt tuningを提案している。これは凍結した汎用大規模言語モデルのプロンプト部分のみを学習する手法である。また [Zhou et al. 23] は通常ならばシステム設計者の試行錯誤で作成されるプロンプトテンプレートについて、遺伝的アルゴリズムで逐次的にプロンプトを改善し、人間である専門家が作成したのと同等の精度を示すプロンプトを自動的に発見する仕組みを提案している。我が国外交実務に際し、大規模言語モデルを実装する際にはこうしたプロンプトを巡る提案を一つないし複数選択し、同時に実装することで類似度という指標を一つとっても本論文で提示した実験結果より良い結果を出すことが可能であるものと考えられる。

[Ouyang et al. 22] は自然言語で記述されたプロンプトに応じて、明示的に学習していないタスクに汎化できる汎用言語モデルの発展について述べるものである。その延長線上で [Wei et al. 21] はinstruction tuningを提案している。これは汎用言語モデルを学習するための手段であり、事前学習済み言語モデルを、様々なタスクについてその指示や例示などのプロンプトとともに学習する。この様にプロンプトを加えることで、学習に含まれていないタスクを解くことが可能になると報告されており、その規模的拡大が試みられるなど [Chung et al. 22]、最近盛んに研究が進められている手法である [Zhang et al. 23] [磯沼ら 23]。一般的なfine tuningではあるタスクに絞って学習が行われるわけだが、これでは他のタスクでは性能が下がってしまう可能性を排除出来ないという問題があった。これに対してinstruction tuningでは学習データにない未知のタ

スクが与えられた時も、タスクを示す何らかの指示を与えることで、その指示に従ってタスクを実行できるように促す。この時、instruction tuningに必要なデータはオープンソースから入手することでコストを下げる事が出来るが、こうしたデータの中に意図的にモデルの予測を大幅に狂わせるデータ (poison data) を混入させることで、大規模言語モデルの性能に負の影響を与えることが可能である点が指摘されている [Wan et al. 23]。こうした効果はたとえ混入するデータの量が少なくても生じる可能性があるものと考えられており、国内外の敵対勢力による介入を絶対的に排除しなければならない我が国の外交実務上、大規模言語モデルを利用する際に考慮すべき重大な論点である。

また、大規模言語モデルについては依然としてその実像が仔細に判明していない点も留意しておく必要がある。本論文では要約作成をタスクとして設定したが、その際、果たして使用した大規模言語モデルであるChatGPTが当該外交文書の全体を通じてその意味から要約を作成しているのか、あるいは複数の部分をいわば抜き書きする形で要約をしているのかが判然としなかったことは事実である。この関連で [Liu et al. 23] は大規模言語モデルが長いコンテキストをどの様に利用しているのかを検証している。そして実験の結果、関連する情報が入力コンテキストの先頭または末尾に在る場合には大規模言語モデルによる計算は最も性能が高くなる事が判明し、他方でこうした情報が入力コンテキストの中間にある場合には同性能が著しく低下することが判明したとしている。本論文において示した実験の結果、どうしても埋めることの出来なかったBERTScore上の類似度の「溝」の部分は、一つにはこうした実験を通じて炙り出される大規模言語モデルの特徴によって生じていたとも考えられよう。また [Chen et al. 23] はそもそもChatGPの性能が短期間で劇的に変化する可能性があることを検証した研究である。そして同研究は大規模言語モデルの性能が中長期的に変化していくことを理解するためにも、継続的な調査が必要であるともしている。ChatGPTに関してOpenAI社が特にGPT-4について技術情報をほぼ開示していないことは、既に触れたとおりである。その意味において大規模言語モデルの我が国外交実務への実装に際しては、その開発各社が性能改善のため施策をとったとしても、関連する情報公開の度合いを常に監視し、最適化を常に図らなければならないのである。

7. 総括と今後の課題

本論文では我が国の外交実務において大規模言語モデルを実装し、行政評価における基準の一つである効率性の観点で有意な改善が見られるかどうかを検証した。大規模言語モデルにはChatGPT (GPT-4) を直接使い、かつ知見と経験の乏しい「若手の末端担当職員」による操作を想定しつつ、入力するプロンプトについてもあえて複雑な工夫を施すことなく、外交実務上、しばしば行われる英文と和文テキストそれぞれについて要約作成を行った。この時、我が国外交実務上は長文の英文テキストにつき「和文」の要

約を作成するのが常であるが、本論文における実験では要約作成を英文であえて行い、翻訳の巧拙は問わないことにした（和文テキストについては和文で要約作成）。英文と和文のテキスト双方について、BERTScoreを評価指標として用い、生成文と参照文の類似度を計測したが、そこでの値をベースにして考えるとこの様に大規模言語モデルそのものを操作して要約作成を行う末端担当職員だけで作業が完結することは想定出来ず、最終的にはその上長、具体的には中間管理職に位置する人物（我が国外務省における首席事務官、総務班長等）によるチェックが必須となることを確認出来た。

このことを踏まえると大規模言語モデルによる外交文書の要約作成を我が国の外交実務において本格的に行うとするならば一方で（多くの場合は若手である）末端担当職員については投入する時間に対する成果物の量という意味での効率性は向上するが、他方で結果として未完成な成果物（要約テキスト案）が大量に作成され、これに対して細かなチェックを最終的に行わなければならないという意味で中間管理職らの作業量は単位時間あたりでむしろ増えてしまう危険性があると言わざるを得ない。しかもこの様に大規模言語モデルを用いた外交実務が行われる様になると末端担当職員についてプロンプトの最適化といった方面での技能こそ向上すれど、最終的に望ましい成果物（本研究でいうならば参照文との関係で類似度の高さを伴う要約文（生成文））を創り出す能力の養成にそれが直結するか否かは定かではないのである。そうした中で若手であった末端担当職員はやがて未熟練なまま外務官僚制の中で昇格し、それまではチェック役を務めてくれていた中間管理職らも昇格し、退職へと至ることになる。すなわち、そこでは暗黙知が形式知として可視化され、確実に伝承されるわけではないのである。大規模言語モデルの外交実務での実装が中長期的な視点で見ると、我が国の外交力そのものの向上をもたらすのかはその意味で未知と言わざるを得ないと考えられる。

本論文の執筆中である2023年11月6日（米東部時間）、OpenAI社は同社公式ホームページにおけるブログという形でGPT-4 Turbo等のリリースを発表した [OpenAI 23]。これはOpenAIの開発者向けカンファレンスであるDevDayの実施と共に行われた発表であり、より安価で、かつより長い文字列（128000トークン）を取り扱うことの出来るGPT-4 Turboの利用を掲げたことでグローバル社会全体より耳目を集めた。本論文の実験では和文テキスト（『外交青書』）の要約を作成する際、ChatGPT（GPT4バージョン）を用いても短時間では処理出来なかった。これに対して新たにリリースされたGPT-4 Turboを用いることで、10万文字以上の和文テキストにつき要約を作成し、可読性のあるテキスト生成に成功したとの公開情報上での簡易的な報告も既に行われている [Classmethod 23]。今後の課題としてはこの様に目覚ましいが不透明な発展を遂げる大規模言語モデルの外交実務への実装により、我が国外交そのものの効率性の向上に資することになるのか、あるいはむしろそこでの実務作業の不安定化要因として効率性を阻害することになるのかについての考察が残されている。

また本論文では初期の考察及び実験に止め、外交実務に必須な秘密保全の視点を踏まえての作業は次なる課題としてとらえることとした。ChatGPTを用いるとしてもより秘密保全の観点から好ましいとされるMicrosoft社のAzure OpenAI Serviceを利用したの考察・実験もその中に含まれる。この点は我が国政府による検討の方向性との関係でも今後留意すべき論点である [デジタル庁 23]。

大規模言語モデルは言うまでもなくOpenAI社によるChatGPTに限られるものではない。我が国においても様々な民間企業等が大規模言語モデルの開発を競って行っており、国立研究開発法人情報通信研究機構（NICT）が総務省と共に整備する学習用言語データはこうした我が国における研究開発のためにアクセス提供されることとなっている [内閣府 23]。今後の課題としては、我が国外交実務への実装に際し、とりわけ秘密保全の観点より、こうした国産の大規模言語モデルを用いる可能性、有効性及び妥当性についての考察・実験が、本論文でも言及した大規模言語モデルの性能向上のための手法をも踏まえつつ、行われるべきであると考えられる。

参考文献

- [相澤 23] 相澤彰子：自然言語処理による学術論文の解析：現状と展望、人工知能 38 (3), pp. 375 - 383 (2023).
- [秋吉ら 22] 秋吉貴雄, 伊藤修一郎, 北山俊哉：公共政策学の基礎 (第3版), 有斐閣 (2022).
- [荒木 23] 荒木千帆美：令和5 年度 (2023 年度) 外務省予算の概要 - 国際秩序の維持・発展に向けた外交実施体制の強化 -, 立法と調査 453, pp. 59 - 65 (2023).
- [安藤 23] 安藤敏行：大規模言語モデルによる特許調査の効率化検討 特許文書の類似文書検索、分類、俯瞰可視化検討, 第20回情報プロフェッショナルシンポジウム予稿集, pp. 43 -48 (2023).
- [Auriemma et al. 23] Auriemma, Serena, et al. "BureauBERTo: adapting UmbERTo to the Italian bureaucratic language." (2023).
- [Beltagy et al. 20] Beltagy, Iz, Matthew E. Peters, and Arman Cohan. "Longformer: The long-document transformer." arXiv preprint arXiv:2004.05150 (2020).
- [Brynjolfsson et al. 23] Brynjolfsson, Erik, et al. "Generative AI at Work." No. w31161. National Bureau of Economic Research (2023).
- [Chen et al. 23] Chen, Lingjiao, "How Is ChatGPT's Behavior Changing over Time?" arXiv preprint arXiv:2307.09009 (2023).
- [Chung et al. 22] Chung, Hyung Won, et al. "Scaling instruction-finetuned language models." arXiv preprint arXiv:2210.11416 (2022).
- [Classmethod 23] Classmethod, "gpt-4-turboのAPIがリリースされたので、10万文字の文章で早速試してみる。Developers IO produced by classmethod" (<https://dev.classmethod.jp/articles/gpt-4-turbo-api-long-tokens/> 2023年11月9日最終閲覧)
- [デジタル庁 23] デジタル社会推進会議幹事会：ChatGPT 等の生成 AI の業務利用に関する申合せ (第2 版) (https://www.digital.go.jp/assets/contents/node/basic_page/field_ref_resources/c64badc7-6f43-406a-b6ed-63f91e0bc7cf/e2fe5e16/20230915_meeting_executive_outline_03.pdf 2023年11月9日最終閲覧)
- [Devlin et al. 18] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

- [福元 20] 福元健太郎：政治学における人工知能の応用に向けて，人工知能 35-4, pp. 526-533 (2020).
- [外務省 18] 外務省：外務省における政策評価の基本計画 (2018).
(https://www.mofa.go.jp/mofaj/annai/page25_001230.html 2023年10月12日最終閲覧)
- [外務人事審議会 21] 外務人事審議会：ポスト・コロナに向けた外交実施体制の抜本的強化：外交・領事実施体制のデジタル・トランフォーメーション (DX) の一層の促進に向けた勧告 (2019).
- [外務省 10] 外務省：外交青書 (2010).
- [外務省 11] 外務省：外交青書 (2011).
- [外務省 12] 外務省：外交青書 (2012).
- [外務省 20] 外務省：外務省デジタル・ガバメント中長期計画
(<https://warp.ndl.go.jp/info:ndljp/pid/12187388/www.kantei.go.jp/jp/singi/it2/cio/siryous3/15gaimu/honbun.pdf> 2023年10月19日最終閲覧)
- [外務省 23] 外務省：外交青書 (2023).
- [原田 23] 原田武夫：自然言語処理技術を用いた我が国外交文書のトピック分析とその外交実務上の活用，立教大学大学院人工知能科学研究科提出修士論文 (2023).
- [He et al. 20] He, Pengcheng, et al. “Deberta: Decoding-enhanced bert with disentangled attention.” arXiv preprint arXiv:2006.03654 (2020).
- [樋口ら 21] 樋口善彦、姉崎正治、糟谷晃弘、小島彰：産業界における技術・技能伝承について，産業技術短期大学誌, 43, pp.29-34 (2021).
- [Hong et al. 22] Hong, Zhi, et al. “Scholarbert: Bigger is not always better.” arXiv preprint arXiv:2205.11342 (2022).
- [磯沼ら 23] 磯沼大、森純一郎、坂田一郎：汎用言語モデル学習のためのプロンプト最適化，人工知能学会全国大会論文集 第 37 回 (2023).
- [伊藤ら 23] 伊藤孝佑，久慈渉，後藤昌夫：特許庁におけるAI技術の活用の現状と最新の取組，情報の科学と技術 (73) 7, pp. 256 -261 (2023).
- [Lan et al. 19] Lan, Zhenzhong, et al. “Albert: A lite bert for self-supervised learning of language representations.” arXiv preprint arXiv:1909.11942 (2019).
- [Lester et al. 21] Lester, Brian et al., “The Power of Scale for Parameter-Efficient Prompt Tuning.” arXiv preprint arXiv:2104.08691 (2021).
- [Lin 04] Lin, Chin-Yew. “Rouge: A package for automatic evaluation of summaries.” Text summarization branches out (2004).
- [Liu et al. 23] Liu, Nelson F. “Lost in the Middle: How Language Models Use Long Contexts.” arXiv preprint arXiv:2307.03172 (2023).
- [Liu et al. 19] Liu, Yinhan, et al. “Roberta: A robustly optimized bert pretraining approach.” arXiv preprint arXiv:1907.11692 (2019).
- [内閣府 23] 内閣府：総務省・NICTが整備する学習用言語データのアクセス提供について
(https://www8.cao.go.jp/cstp/ai/ai_senryaku/5kai/datateikyuu.pdf 2023年11月9日最終閲覧)
- [難波 20] 難波英嗣：テキスト間の類似度の測定，情報の科学と技術 (70) 7, pp. 373 -375 (2020).
- [日本放送協会 23] 日本放送協会 (NHK), 「ChatGPT」開発企業のアルトマンCEO 岸田首相と面会
(<https://www3.nhk.or.jp/news/html/20230410/k10014034131000.html> 2023年10月19日最終閲覧)
- [Noy et al. 23] Noy, Shakked, “Eperimental Evidence on the Productivity Effects of Generative Artificial Intelligence.” (Available at SSRN 4375283) (2023).
- [OECD 23a] Organization for Economic Cooperation and Development “REVIEW OF THE OECD RECOMMENDATION ON CROSS-BORDER CO-OPERATION IN THE ENFORCEMENT OF LAWS PROTECTING PRIVACY “ (2023).
- [OECD 23b] Organization for Economic Cooperation and Development “INITIAL POLICY CONSIDERATIONS FOR GENERATIVE ARTIFICIAL INTELLIGENCE” (2023).
- [OECD 23c] Organization for Economic Cooperation and Development “ENHANCING THE SECURITY OF COMMUNICATION INFRASTRUCTURE” (2023).
- [OECD 23d] Organization for Economic Cooperation and Development “G7 HIROSHIMA PROCESS ON GENERATIVE ARTIFICIAL INTELLIGENCE (AI), TOWARDS A G7 COMMON

- UNDERSTANDING ON GENERATIVE AI" (2023).
- [OECD 23e] Organization for Economic Cooperation and Development "TRANSPARENCY REPORTING ON CHILD SEXUAL EXPLOITATION AND ABUSE ONLINE" (2023).
- [OECD 23f] Organization for Economic Cooperation and Development "REGULATORY SANDBOXES IN ARTIFICIAL INTELLIGENCE" (2023).
- [OECD 23g] Organization for Economic Cooperation and Development "ONLINE PRODUCT SAFETY SWEEP REPORT" (2023).
- [OECD 23h] Organization for Economic Cooperation and Development "MOVING FORWARD ON DATA FREE FLOW WITH TRUST. NEW EVIDENCE AND ANALYSIS OF BUSINESS EXPERIENCES" (2023).
- [OECD 23i] Organization for Economic Cooperation and Development "AI LANGUAGE MODELS TECHNOLOGICAL, SOCIO-ECONOMIC AND POLICY CONSIDERATIONS" (2023).[OECD 2023j] Organization for Economic Cooperation and Development "EMERGING PRIVACY ENHANCING TECHNOLOGIES. CURRENT REGULATORY AND POLICY APPROACHES" (2023).
- [OECD 23j] Organization for Economic Cooperation and Development "Tracing the impacts of Russia's war of aggression against Ukraine on official development assistance (ODA)" (2023).
- [OECD 23k] Organization for Economic Cooperation and Development "What are the integration challenges of Ukrainian refugee women?" (2023).
- [OECD 23l] Organization for Economic Cooperation and Development "Recovering lost learning opportunities in Ukraine: Key education policy strategies" (2023).
- [OECD 23m] Organization for Economic Cooperation and Development "Confronting the energy crisis: Changing behaviours to reduce energy consumption" (2023).
- [OECD 23n] Organization for Economic Cooperation and Development "Responsible business conduct implications of Russia's invasion of Ukraine" (2023).
- [OECD 23o] Organization for Economic Cooperation and Development "GLOBAL VALUE CHAIN DEPENDENCIES UNDER THE MAGNIFYING GLASS" (2023).
- [OECD 23p] Organization for Economic Cooperation and Development "CHALLENGES TO INTERNATIONAL TRADE AND THE GLOBAL ECONOMY RECOVERY FROM COVID-19 AND RUSSIA'S WAR OF AGGRESSION AGAINST UKRAINE" (2023).
- [OECD 23q] Organization for Economic Cooperation and Development "What we know about the skills and early labour market outcomes of refugees from Ukraine" (2023).
- [OECD 23r] Organization for Economic Cooperation and Development "The intergovernmental fiscal outlook and the implications of Russia's war against Ukraine, high energy prices and inflation" (2023).
- [OECD 23s] Organization for Economic Cooperation and Development "Turning to regions and local governments to rebuild Ukraine" (2023).
- [OpenAI 23a] OpenAI "GPT-4 Technical Report" (2023)
(<https://cdn.openai.com/papers/gpt-4.pdf> 2023年10月19日最終閲覧)
- [OpenAI 23b] OpenAI "New models and developer products announced at DevDay. GPT-4 Turbo with 128K context and lower prices, the new Assistants API, GPT-4 Turbo with Vision, DALL-E 3 API, and more."
(<https://openai.com/blog/new-models-and-developer-products-announced-at-devday> 2023年11月9日最終閲覧)
- [Ouyang et al. 22] Ouyang, Long, et al. "Training language models to follow instructions with human feedback." *Advances in Neural Information Processing Systems* 35, ppt27730-27744 (2022).
- [Papineni et al. 02] Papineni, Kishore, et al. "Bleu: a method for automatic evaluation of machine translation." *Proceedings of the 40th annual meeting of the Association for Computational Linguistics* (2002).

- [Paulus et al. 17] Paulus, Romain, Caiming Xiong, and Richard Socher. "A deep reinforced model for abstractive summarization." arXiv preprint arXiv:1705.04304 (2017).
- [Reimers et al. 19] Reimers, Nils, and Iryna Gurevych. "Sentence-bert: Sentence embeddings using siamese bert-networks." arXiv preprint arXiv:1908.10084 (2019).
- [Reiter 18] Reiter, Ehud. "A structured review of the validity of BLEU." Computational Linguistics 44.3, pp393-401 (2018).
- [蒔田 20] 蒔田純：政策形成におけるAI活用に関する一考察（上），弘前大学教育学部紀要 124, pp31-40 (2020).
- [西尾 01] 西尾勝：行政学（新版），有斐閣（2001）。
- [野中・竹内 20] 野中郁次郎，竹内弘高：知識創造企業（新装版），東洋経済新報社（2020）。
- [Radford et al., 18] Radford, Alec, et al. "Improving language understanding by generative pre-training." (2018).
- [竹本 11] 竹本信介：戦後日本における外務官僚のキャリアパス —誰が幹部になるのか？—，立命館法学337, pp.1573-1686 (2011)。
- [田中ら 23] 田中大智，重松祐匡，木山真人，尼崎太樹：文章の意味的類似指標に関する一検討，人工知能学会全国大会論文集 第 37 回 (2023)。
- [Vaswani et al. 17] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).
- [Von Hippel 94] Von Hippel, Eric : "'Sticky information" and the locus of problem solving : implications for innovation", Management Science 40. 4, pp.429-439 (1994).
- [Wan et al. 23] Wan, Alexander, et al. "Poisoning Language Models During Instruction Tuning." arXiv preprint arXiv:2305.00944 (2023).
- [Wei et al. 21] Wei, Jason, et al. "Finetuned language models are zero-shot learners." arXiv preprint arXiv:2109.01652 (2021).
- [山田ら 23] 山田育矢，鈴木正敏，山田康輔，李凌寒：大規模言語モデル入門，技術評論社（2023）。
- [山下・赤井 05] 山下耕治，赤井伸郎：財政運営における透明性（情報公開，政策評価）と事業効率性，長崎大学経済学部研究年報 21, pp.61-73 (2005)。
- [吉原・山谷 09] 吉原健吾，山谷清志：外務省における政策評価：「政策評価」と「外交」評価の交錯，同支社政策研究 = Doshisha University Policy Studies 3, pp.141-156 (2009)。
- [Zhang et al. 18] Zhang, Tianyi, et al. "Bertscore: Evaluating text generation with bert." arXiv preprint arXiv:1904.09675 (2019).
- [Zhang et al. 23] Zhang, Shengyu, et al. "Instruction tuning for large language models: A survey." arXiv preprint arXiv:2308.10792 (2023).
- [Zhou et al. 23] Zhou, Yongchao et al. "Large Language Models Are Human-level Prompt Engineers", arXiv preprint arXiv:2211.01910 (2022).

(本学非常勤講師)

