

# 大規模言語モデルを用いた 可読性評価指標の提案: 産業日本語適合度

Proposal for a Readability Index Using Large Language Model: Technical Japanese Conformity



株式会社日本取引所グループ 総合企画部 主任研究員

## 土井 惟成

2015年 株式会社日本取引所グループに入社。東京証券取引所 IT 開発部などを経て、2020年より現職。  
2022年より株式会社JPX 総研インデックスビジネス部を兼務

✉ n-doi@jpx.co.jp

### 1 はじめに

証券市場は、様々な投資家がそれぞれの投資判断に基づいて株式などの売買を行う場であり、上場会社が開示する情報がその判断の重要な基盤となっている。これらの開示情報の一つに、金融商品取引法に基づく有価証券報告書がある。上場会社は、企業の概況や経理の状況等を記載した有価証券報告書を、事業年度末から三ヶ月以内に内閣総理大臣に提出することが義務付けられている。

有価証券報告書の記載事項の一つに、「経営者による財政状態、経営成績及びキャッシュ・フローの状況の分析」(Management's Discussion and Analysis of Financial Condition and Results of Operations、以下、MD&A)がある。MD&Aの開示により、投資家は、「企業が経営方針・経営戦略等を遂行するに当たっての財源の十分性」や「企業の経営方針・経営戦略等の実現可能性」の判断が可能となる<sup>[1]</sup>。従って、有価証券報告書におけるMD&Aは、重要な定性情報源として認識されている<sup>[2]</sup>。

金融領域におけるテキスト分析では、テキストの素性として可読性に着目し、経済的影響を検証する研究がある<sup>[3]</sup>。廣瀬ら<sup>[4]</sup>は、MD&Aの可読性と企業の将来の業績との関連性を指摘し、MD&Aのテキストの可読性と、時価総額をはじめとする主要な財務指標との相関関係を明らかにした。ここで使用された可読性の指標は、柴崎・玉岡<sup>[5]</sup>による、以下の多変量回帰式である。

$$Y = -0.145X_1 + 0.587X_2 + 14.016 \quad (1)$$

ここで、 $X_1$ は文の総文字数に対する平仮名の割合(%)を示し、 $X_2$ は平均述語数を示す。 $Y$ が小さいほど、その文の可読性が高いことを示す。しかしながら、式(1)を用いた可読性の評価は、内容の理解に必要なリテラシーの評価指標を算出するものであり、産業文書における情報の理解しやすさを測定できるとは限らない。

金融領域のテキストの分析に、大規模言語モデルの活用が進んでいる。Kuroki et al.<sup>[6]</sup>は、決算説明会のスクリプトからの主観的表現の抽出に、Doi et al.<sup>[7]</sup>は、監査上の主要な検討事項のトピック分類に、それぞれ大規模言語モデルを利用している。また、MD&Aの分析に大規模言語モデルを利用した報告として、センチメントに関するスコアの付与<sup>[8]</sup>、MD&Aの要約文を生成し、それを基にした情報の密度に関する指標の提案<sup>[9]</sup>、そしてMD&Aの記述内容における主観的表現の抽出<sup>[10]</sup>が挙げられる。このように、金融領域のテキストにおける大規模言語モデルの有用性を踏まえると、産業日本語の解析にも大規模言語モデルが有用である可能性がある。

そこで、新たなテキストの可読性の評価指標として、大規模言語モデルを利用して算出する、「産業日本語適合度」を提案する。産業日本語とは、「産業・技術情報を人に理解しやすく、かつ、コンピュータ(機械)にも処理しやすく表現するための日本語」と、産業日本語研究会によって定義されている。大規模言語モデルに対して、入力文を産業日本語に変換するよう指示し、その出力文と入力文の類似度を、「産業日本語適合度」と定義する。産業日本語適合度は、入力文の「産業日本語らしさ」を表し、高いほど産業日本語らしい文と言える。

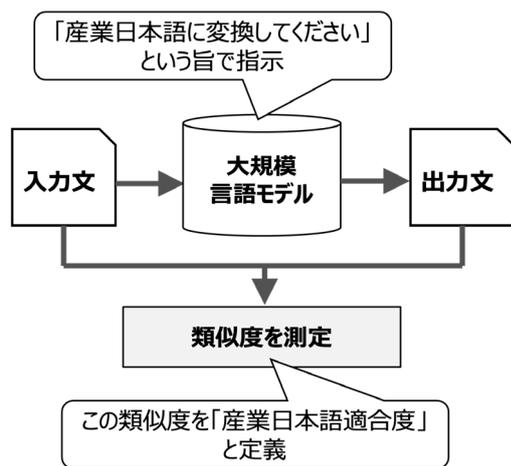


図1 提案手法の概要

提案手法の概要を図1に示す。

本研究では、産業日本語適合度の算出方法を提案し、その実証研究を行う。具体的には、上場会社のMD&Aと、特許出願の要約書における概要（以下、特許の概要）及び特許請求の範囲における最初の請求項（以下、第一請求項）について、産業日本語適合度を測定する。そして、人手による可読性評価結果との相関を分析し、また、文書の種類別の産業日本語適合度の傾向を確認する。

本稿の構成は次の通りである。第2章では、テキストの可読性に関する関連研究を述べる。第3章では、産業日本語適合度の算出方法を提案する。第4章では、実際に使用するデータセットを述べる。第5章では、産業日本語適合度と人手評価との相関に関する実験と、文書の種類別の産業日本語適合度の傾向を把握するための実験について述べる。最後に、第6章で、結論をまとめる。

## 2 関連研究

テキストの可読性は、自然言語処理システムの出力結果の品質評価をはじめとして、幅広い用途で利用されている<sup>[11]</sup>。その中でも、内容の理解に必要なリテラシー（学校教育年数）を表したものが、可読性の評価指標として広く利用されている。日本語を対象とした同様の評価手法としては、日本語の教科書コーパスに基づいて導出された、1文字ごとの集計値を用いたSato et al.の評価式<sup>[12]</sup>や、形態素等を用いた柴崎・玉岡の評価式<sup>[15]</sup>が挙げられる。また、英語とスロベニア語を対象とした研究では、ディープニューラルネットワークを用いた可読性の評価手法も提案されている<sup>[13]</sup>。

これらの評価手法はいずれも、内容の理解に必要なリテラシーの指標を算出するものであり、「テキストに含まれる産業・技術情報の理解しやすさ」と必ずしも関連するとは限らない。すなわち、従来手法では、産業日本語らしさを測定できるとは必ずしも限らないものと考えられる。

## 3 提案手法

本研究では、大規模言語モデルを用いたテキストの可読性の指標として、産業日本語適合度を提案する。

大規模言語モデルに入力するプロンプトの例を、下記のとおり示す。このプロンプトでは、前文に産業日本語への変換を指示する文を入力し、その後に対象の文を入力する。また、プロンプトのフォーマットは、マークダウン形式のテキストを意識した。これにより、大規模言語モデルに対して、より明確に指示が伝わるものと考えられる。

下記の入力を、「産業・技術情報を人に理解しやすく、かつ、コンピュータ（機械）にも処理しやすく表現するための日本語」に変換し、その文を出力してください。ただし、出力には、入力における重要情報の全てを、正確に含んでください。

```
## 入力
...
## 出力
```

なお、大規模言語モデルは、必ずしも産業日本語の概念を理解しているとは限らない可能性がある。そのため、産業日本語への変換の指示においては、「産業日本語」という単語を用いず、その定義をプロンプトに明記した。

また、大規模言語モデルによる言い換えの過程で、入力文の重要情報が欠落する可能性がある。これを回避するため、「特許文献機械翻訳の品質評価手順」<sup>[14]</sup>の「内容の伝達レベルの評価」の表現を参考とし、入力文の重要情報を欠落させない旨を指示をプロンプトに含めた。

本研究では、プロンプトにおける指示は簡潔に留め、「特許ライティングマニュアル」<sup>[15]</sup>の言い換えルールは参照しないこととした。この理由として、当該マニュアルの言い換えルールは、具体的な文法や表現に関する指示が多く、変換の際にノイズとなる可能性が懸念され

たことが挙げられる。その代わりに、本研究では、簡潔な指示文を用いることで、大規模言語モデルの柔軟性を活かし、入力文の産業日本語への変換をより自然に行えるようにした。

類似度の評価指標には、BLEU (BiLingual Evaluation Understudy)<sup>[16]</sup>を採用した。BLEUスコアは、元々は機械翻訳の評価に用いられており、ある文書の単語列が別の文書の単語列とどの程度一致するかを測定し、n-gram (連続するn個の単語の列)の一致率を基に計算される。BLEUスコアは、算出方法の特性上、文脈や意味の類似性は直接反映されないという限界がある。しかしながら、本研究では、出力文と入力文の単語単位での類似性の評価に焦点を当てることとし、BLEUスコアを採用した。

また、大規模言語モデルの出力は、入力ごとにランダムに変化する。そこで、本手法では、同一のプロンプトを大規模言語モデルに10回入力し、それぞれについてBLEUスコアを算出し、それらの中央値(5番目と6番目に大きいBLEUスコアの平均値)を産業日本語適合度とした。

## 4 データセット

本研究では、提案手法である産業日本語適合度の有用性を検証するために、産業に関する特定のテキストを使用して評価実験を行う。本章では、後述する実験で使用するデータセットについて詳述する。

本研究で使用するデータセットのデータソースとして、上場会社の有価証券報告書のMD&Aセクションと、特許の概要と第一請求項を対象とした。これらのテキストは、実務的に使用されているテキストであり、産業日本語で記載されることが望ましいものと考えられる。

本研究では、これらのデータソースから、200字から400字の範囲のテキストをランダムに100件ずつ抽出し、後述の実験で使用する。これらのデータソースにおいては、文字列長の分布が異なる可能性がある。特に、特許の概要は、その上限が400字と定められており、他のテキストよりも短い可能性がある。このような分布の違いが、評価実験に影響を及ぼすことを避けるため、本研究では実験の対象とするテキストの文字列長について、200字から400字の範囲という制限を設けた。

MD&Aのデータセットの構築においては、2023年10月末時点でTOPIX 500に選定されている上場会社を対象として、2022年7月1日から2023年6月30日までの期間に開示された、2022年4月1日から2023年3月31日を期末とする有価証券報告書を使用した。そして、HTML形式の有価証券報告書から、MD&Aのセクションを抽出し、その本文の中からpタグごとにテキストを抽出することで、パラグラフ単位のテキストを抽出した。この結果、45,796件のテキストを抽出し、この中から200字以上400字以下のテキストを3,554件抽出し、その中からランダムに100件のテキストを選択した。

特許の概要及び第一請求項のデータセットの構築においては、2022年6月に登録公報(B2)が発行された特許を対象とした。また、典型的な特許を対象とするため、概要に【課題】と【解決手段】を含む特許を抽出し、その中から、特許の概要と第一請求項がそれぞれ200字以上400字以下の特許を抽出した。この結果、全体で14,270件の特許から、3,942件の特許を抽出し、この中からランダムに100件の特許を選択した。この抽出した特許から、特許の概要と第一請求項を抽出し、データセットを構築した。

以上により、1つのテキスト当たりの文字列長が200字以上400字以下の、MD&A、特許の概要及び第一請求項のデータセットを、100件ずつ作成した。

## 5 実験

本研究では、第4章で構築したデータセットを用いて、産業日本語適合度を算出した後、2種類の実験を行う。

まず、人手評価との相関分析を行う。この時、柴崎・玉岡<sup>[5]</sup>による可読性(以下、式(1))との相関分析も併せて行うことで、提案手法の有用性を評価する。

次に、各データセットごとに産業日本語適合度の分布を確認する。これにより、文書の種類ごとの産業日本語適合度の傾向を把握する。

提案手法による産業日本語適合度の算出において、大規模言語モデルとしては、本稿時点におけるOpenAI社の最新のChatGPTのモデルである、gpt-4o-2024-05-13(以下、GPT-4o)を採用した。日本語の形態素解析にはMeCab<sup>[17]</sup>とUniDic<sup>[18]</sup>を使用し

た。BLEU スコアの算出には、SacreBLEU<sup>[19]</sup> を使用し、n-gram の n の値は、デフォルト値である 4 とした。

## 5.1 人手評価との相関分析

まず、各データセットのテキストに対して、人手で可読性の評価を行った。この評価では、著者が、「特許ライティングマニュアル」<sup>[15]</sup> を参照し、対象のテキストが十分に産業日本語であれば 1 を付与し、産業日本語とするためには一定の編集が必要な場合には 0 を付与した。なお、同マニュアルでは、言い換えルールとして「短文化」が定められており、その言い換えを行うべき基準は 100 文字とされている。しかしながら、本データセットにおけるテキストの多くは、含まれている文の文字列長が 1 文当たり 100 文字を大幅に超過しているながらも、読みやすさが損なわれていない傾向がある。そこで、本評価では、テキストに含まれている文が 1 文当たり 200 文字を超過している時は、短文化を積極的に検討するようにしつつも、定量的な条件は設けないこととした。この結果、各データセットにおける 1（十分に産業日本語である）の文の分布は、MD&A が 75 件、特許の概要が 46 件、第一請求項が 22 件となった。

合計 300 件のテキストに対する人手評価について、式 (1) による可読性と産業日本語適合度のそれぞれとピアソンの積率相関係数を算出した。この結果、式 (1) による可読性との相関係数は -0.446、産業日本語適合度との相関係数は 0.496 となった。これらの絶対値を比較すると、産業日本語適合度の方が大きいことから、提案手法の有用性が示された。従って、テキストの産業日本語適合度が、実際の産業文書における情報の処理や理解の容易さをより直接的に反映していると言える。

産業日本語適合度の方が相関係数の絶対値が大きくなった理由として、式 (1) が、深い文脈的な理解や、複雑な文法構造を考慮していないことが挙げられる。MD&A や特許文書をはじめとする産業で使用されるテキストでは、複雑な表現を有していることが多くあり、式 (1) ではこのような表現の文において可読性が低くなるようにスコアを算出できなかった可能性がある。これに対して、GPT-4o は、テキストが持つ潜在的な特性や文脈を捉えながら入力文の産業日本語への変換を実現できたため、結果的に入力文がどの程度産業日本語の基準に適合しているかを定量的に評価できたと考えられる。

ただし、これらの相関分析の実験結果の限界として、人手評価の主観性とデータセットの規模の小ささが挙げられる。本評価における人手による可読性の評価は、評価者のバイアスや評価基準の一貫性の欠如により、結果に一定の偏りが生じている可能性がある。更に、今回の実験では、評価者は著者一人のみであり、データセットの規模も合計 300 件と限られている。そのため、今後の分析においては、統計的な信頼性が確保されるほどの大きなデータセットでの検証が必要であると言える。また、今後のデータセットの構築においては、複数人による評価が必要であり、その評価における人手評価の主観性を軽減するためには、評価基準の明確化が必要である。

## 5.2 産業日本語適合度の分布の分析

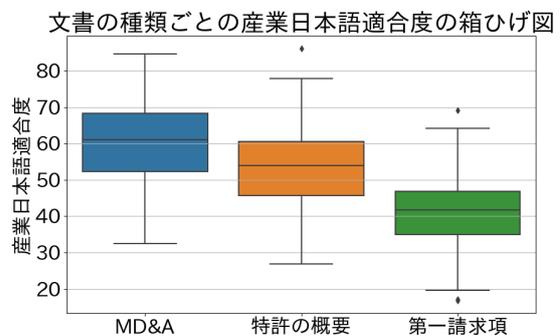


図 2 産業日本語適合度の分布

本節では、第 4 章で説明したデータセットを用いて、産業日本語適合度の分布を分析し、文書の種類ごとに産業日本語適合度の分布を確認する。

文書の種類ごとの産業日本語適合度の分布を図 2 に示し、各文書種類の GPT-4o の出力の例を表 1 に示す。

各文書種類の産業日本語適合度の平均値を算出すると、MD&A は 60.150、特許の概要は 54.072、第一請求項は 41.246 であった。また、t 検定の結果、MD&A と特許の概要、及び特許の概要と第一請求項の間での産業日本語適合度の平均値には、それぞれ統計的に有意な差が認められた ( $p < 0.01$ )。これらの結果から、MD&A、特許の概要、第一請求項の順で、産業日本語適合度が大きいことが示された。

このような産業日本語適合度の差異は、それぞれのテキストにおける内容の伝達のしやすさと流暢さの傾向に起因しているものと考えられる。特に、第一請求項は、専門性の高い用語が使用されるうえに、特有の文体で記



表1 産業日本語適合度の算出例

#	種類	入力	出力の例	産業日本語適合度	出典
1	MD&A	当連結会計年度の製品売上数量は、前期比97.8%となりました。主な要因は、前期のコロナ下での行動制限に伴う内食需要増加の反動等により、生鮮食品向けを中心に出荷量の減少が見られたこと、原料価格の急騰を受けた第二次製品価格改定の交渉を行う中でシェア拡大に向けた活動を一時中断していたことによるものです。なお、コロナ前である2020年3月期比の製品売上数量は108.2%、2020年3月期からの年平均成長率(CAGR)は2.7%となりました。	当連結会計年度の製品売上数量は、前期比97.8%となりました。主な要因は、前期のコロナ下での行動制限に伴う内食需要増加の反動により、生鮮食品向けを中心に出荷量が減少したこと、及び原料価格の急騰を受けた第二次製品価格改定の交渉中にシェア拡大活動を一時中断していたことです。なお、コロナ前である2020年3月期比の製品売上数量は108.2%、2020年3月期からの年平均成長率(CAGR)は2.7%となりました。	82.949	株式会社エフピコ 有価証券報告書 (第61期)
2	特許の概要	【課題】カニューレアセンブリ、および、カニューレアセンブリを製造する方法の提供。 【解決手段】該カニューレアセンブリは、カニューレと、該カニューレに連結されたピグテール延長部を含む。該ピグテール延長部は、第一の剛性を有する近位部と、第二の剛性を有する遠位部とを含み、該第一の剛性は該第二の剛性より大きい。該ピグテール延長部の該近位部は、該カニューレと該遠位部の少なくとも一部分との間に位置決めされる。	【課題】カニューレアセンブリとその製造方法の提供。 【解決手段】このカニューレアセンブリは、カニューレと、それに連結されたピグテール延長部を含む。ピグテール延長部は、第一の剛性を持つ近位部と、第二の剛性を持つ遠位部から成り、第一の剛性は第二の剛性より大きい。近位部は、カニューレと遠位部の少なくとも一部分の間に位置する。	49.032	特開 2021-058791
3	第一請求項	電動の回転工具に装着することができるテープ巻取り用具であって、前記電動の回転工具に着脱可能な連結部と、前記連結部から延び、テープを巻取り可能な細い棒状の巻取部と、前記巻取部の長手方向に延びるように設けられ、テープを差し込み可能であるスリット状の開口部と、を備え、前記開口部は、前記巻取部の長手方向における前記連結部とは反対側の先端部に開放せず、前記巻取部を貫通するように形成されるテープ巻取り用具。	電動回転工具に取り付け可能なテープ巻取り用具は、以下の構成要素を持ちます。まず、電動回転工具に着脱できる連結部があります。次に、連結部から延びる細い棒状の巻取部があり、この部分でテープを巻き取ることができます。さらに、巻取部の長手方向に沿ってスリット状の開口部が設けられており、この開口部にテープを差し込むことが可能です。この開口部は、巻取部の長手方向において連結部とは反対側の先端部に開放されず、巻取部を貫通する形で形成されています。	31.760	特許 7085410

載されている傾向があるため、一般的な産業日本語の基準に合わせると産業日本語適合度は低くなりやすいと考えられる。このように、異なる種類の文書において産業日本語適合度の分布が異なることから、産業日本語適合度は、文書ごとの特性の理解に役立つ指標であると考えられる。

ただし、その内容の伝達のしやすさと流暢さの傾向が、産業日本語適合度の大小へどのように影響を与えているかについては、個別の事例や他の分野のテキストを踏まえた調査が必要である

## 6 おわりに

本研究では、大規模言語モデルを活用した新たな可読性の評価指標として、産業日本語適合度を提案し、その有用性の評価実験と文書の種類別の傾向分析を行った。産業日本語適合度は、「産業日本語らしさ」を定量的に評価するものである。実験の結果、産業日本語適合度は、従来の可読性評価手法よりも、人手評価との高い相関を示した。

更に、MD&A、特許の概要及び第一請求項について、産業日本語適合度を測定したところ、各文書の産業日本語適合度に有意な差が存在することが明らかになった。

特に、MD&A が最も高い産業日本語適合度を示し、投資家向け情報としての可読性の高さが確認された。これに対して、特許文書、特に第一請求項は、固有の文体と専門性の高さから、産業日本語適合度が相対的に低くなる傾向にあることが確認された。

今後の研究としては、産業日本語適合度の測定方法をさらに精緻化し、その検証を広範囲にわたるテキストのサンプルで行うことが考えられる。また、今回の実験では、大規模言語モデルとして GPT-4o という、詳細が公開されていないモデルを使用したが、今後においては、LLM-jp-13B v2.0 [20] をはじめとする、日本語に特化したオープンな大規模言語モデルによる実験を行うことが考えられる。

## 参考文献

- [1] 金融庁．記述情報の開示に関する原則，2019.
- [2] 加藤大輔，五島圭一．有価証券報告書のテキスト分析：経営者による将来見通しの開示と将来業績，金融研究，Vol. 40, No. 3, pp. 45-75, 2021.
- [3] 首藤昭信．テキスト分析と会計学研究，情報センサー，Vol. 143, pp. 8-10, 2019.
- [4] 廣瀬喜貴，平井裕久，新井康平．MD&A 情報の可読性が将来業績に及ぼす影響：テキストマイニングによる

- 分析, 経営分析研究, No. 33, pp. 87-101, 2017.
- [5] 柴崎秀子, 玉岡賀津雄. 国語科教科書を基にした小・中学校の文章難易学年判定式の構築, 日本教育工学会論文誌, Vol. 33, No. 4, pp. 449-458, 2010.
- [6] Y. Kuroki, T. Manabe and K. Nakagawa. Fact or Opinion? – Essential Value for Financial Results Briefing. In 2023 14th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), pp. 375-380, 2023.
- [7] N. Doi, Y. Nobuta, and T. Mizuno. Topic Classification of Key Audit Matters in Japanese Audit Reports by Zero-shot Text Classification. In 2023 14th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), pages 540-545, 2023.
- [8] Hu, N., Liang, P., and Yang, X. Whetting All Your Appetites for Financial Tasks with One Meal from GPT? A Breakthrough Comparison of GPT-3, Finbert, and Dictionaries in Sentiment Analysis. SSRN Scholarly Paper, 2023.
- [9] Kim, A. G., Muhn, M., and Nikolaev, V. V. Bloated Disclosures: Can ChatGPT Help Investors Process Information?, Chicago Booth Research Paper, No. 23-07, 2023.
- [10] 屋嘉比潔, 黒木裕鷹, and 中川慧. MD&A における定性的表現と経営者予想の精度, 人工知能学会全国大会論文集, 第 38 回 (2024), 一般社団法人人工知能学会, 2024.
- [11] Saku Sugawara, Yusuke Kido, Hikaru Yokono, and Akiko Aizawa. Evaluation metrics for machine reading comprehension: Prerequisite skills and readability. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 806-817, 2017.
- [12] Satoshi Sato, Suguru Matsuyoshi, and Yohsuke Kondoh. Automatic assessment of Japanese text readability based on a textbook corpus. In Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC' 08), Marrakech, Morocco, European Language Resources Association (ELRA). 2008
- [13] Matej Martinc, Senja Pollak, and Marko Robnik-Šikonja. Supervised and unsupervised neural approaches to text readability. Computational Linguistics, Vol. 47, No. 1, pp. 141-179, 2021.
- [14] 特許庁情報技術統括室. 特許文献機械翻訳の品質評価手順 Ver.1.0, 2014.
- [15] 一般財団法人日本特許情報機構特許情報研究所. 特許ライティングマニュアル「産業日本語」第 2 版. 2019.
- [16] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics, pages 311-318, 2002.
- [17] Taku Kudou. Mecab : Yet another part-of-speech and morphological analyzer. 2005. <http://mecab.sourceforge.net/> (2024.5.31 閲覧)
- [18] 国立国語研究所. 「UniDic」国語研短単位自動解析用辞書. <https://clrd.ninjal.ac.jp/unidic/> (2024.5.31 閲覧)
- [19] Matt Post. A call for clarity in reporting BLEUScores. In Proceedings of the Third Conference on Machine Translation: Research Papers, pages 186-191, 2018.
- [20] 国立情報学研究所. 大規模言語モデル「LLM-jp-13B v2.0」を構築, 2024. [https://www.nii.ac.jp/news/upload/nii\\_newsrelease\\_20240430.pdf](https://www.nii.ac.jp/news/upload/nii_newsrelease_20240430.pdf) (2024.5.31 閲覧)