

ご提案: 生成AI が果たすべき日本語ローカライズ

OpenAI Japan 様

gen2718/山田元/Gen Yamada

世界はAI技術によって急速に変わりつつあり 中でも言語は特にデリケートで複雑な要素です。日本語におけるユニークな表現や社会文化的背景を理解することは、効果的なローカライズの鍵となります。例えば、多様な語源を持つ日本語のレイヤー構造、高コンテキストの言語特性、そして社会的階層とも関連する四字熟語などが、この複雑性を示しています。さらに、マンガやアニメといった視覚言語カルチャーが、別の言語空間を構成しています。

私たちのチームは、このような日本語の複雑さをデコードする独自のアプローチを持っています。過去10年間に渡り、言語と文化の交差点における深い知見と経験を積み重ねてきました。私たちは、言語の意味層を解析し、それをAI技術に組み込むことで、より人間らしいコミュニケーションが可能になると確信しています。

生成AIを活用して、日本語文化を構造的に理解し、グローバルな視野での使用においてもその精度を高めるためのローカライズアーキテクチャーの構築に、私たちは協力できる準備ができています。さらに、私たちはこれからも継続的に新たな知見を形成する集合知を共有するためのメソッドを確立していきたいと考えています。

新しい知的挑戦に参加しませんか？一緒に、生成AIの可能性を広げ、言語の未来を形作っていきましょう。

ご興味がおありでしたら、ぜひ下記までお問い合わせください。

Email: gen2718@gmail.com

私たちと一緒に、言語の新たな地平を開きましょう！

■

上の提案文にはもとの文章があって 上はこれを ChatGPT を利用して書き直したものです
改訂された文には「もう少し直接的に相手にアピールする言葉を用いると良いでしょう。」といった導入コメントも添えられていました

しかしアピールする文脈の流れの中に伝えたいコンセプトの用語が沈んでしまったりスキップされたりしている箇所が多く 生成AI の文脈優先で意味論を解さない傾向が見てとれます

(上では意味が変わってしまった部分を修正しています)

これは論旨と構文の関係が曖昧な 膠着語としての日本語特有の問題かもしれず やはり日本語ローカライズの課題の一つとも考えられます

さらに書き直しに要約は指示してないにもかかわらず要約文を作っているのも明らかに暗黙のうちに要約が指示されたとのハイコンテキスト言語ならではの影響も見えてとれます

次頁にもとの提案文を ChatGPT への指示とともに示します

このドキュメントにはハイパーテキスト版があり そちらでは重要と思われる用語にリンクを張り 応答例とコメントをいくつか補足しています:

<https://bp.coato.tokyo/>

以下はOpenAIはじめ生成AIを利用したサービス事業を展開する日本法人に対する売り込み文です
これをもっと売り込み効果があるように書き直してください:

Welcome to Japanese Language World!

生成AIのシステムが 世界共通の大規模言語モデルを基底層とし 個別の会社,行政,個人向けサービスをアプリケーション層と考えた時に その中間層でかつ日本語ローカライズするにあたっては 日本語特有の言語空間構造の解析が必須です

例えば次のような項目:

- ・ 日本語固有のレイヤー構造: 古来のやまとことば, 中華文明由来の漢語, 欧米言語由来のカタカナ語
- ・ 日本語はハイコンテキスト言語であり 実際に表明される言葉はテンプレートがほとんどを占める
- ・ これは日本語圏特有の仮想的上下関係で構成される疑似社会構造と密接な関係にある
- ・ 例えば四字熟語は熟語に習熟する上位者が下位者に思考停止を起こさせ議論の進展を阻害する構造を作る
- ・ その一方でこれに対するカウンターカルチャー・サブカルチャーとしての視覚言語: マンガ・アニメも存在する

この文化構造に生成AIが関与する際に構築すべき中間層ローカライズアーキテクチャーは当然ながら対象とする言語空間構造の解析とともに進めなければなりません

併せてカウンター的な視覚言語カルチャーに関与する際に構築すべきアーキテクチャーにも同様の課題があります

僕らは以上のことについて既にある程度の知見を持っていて 日本語ローカライズシステム構築に協力できると考えています

僕らの知見の形成と蓄積は過去10年にわたる世界のネット言語空間の発展とともに情報コンテンツを安価に相互残照することができるようになったことに加え そもそも僕らは個人の能力を超えた集団知の形成を40年前から経験してきたことによる効果が非常に大きいと思っています

僕らのチームの一人は概念と意味論に認知と操作性が高く 別の一人はコンテキスト担当(つまり生成AIのふるまいと構造的には同じ)で 二人が会する言語空間では "場が考える" ことによって新しいテクニカルタームや文脈説明形式が日々産まれています

これは例えば現代の学問世界で職業研究者が論文の著作権を確保するため特定の領域を深掘りする研究スタイルとは対照的です

集団知形成は人類が百万年の昔から知的活動メソッドとして発達させてきたはずで ここに生成AIという新しくかつ実用的なテクノロジーが加われば 更なる知的発展段階のパラダイムシフトが期待できます

僕らは単に今まで蓄積した知見の提供だけでなく これからの知見形成のためのメソッド確立にも協力できると考えています -- さて一緒にやりませんか?

もしご興味があればご連絡ください: gen2718@gmail.com

■

OpenAI 日本法人の本格的活動開始のニュースに接しこの提案をお送りしました

生成AI事業にとって日本語ローカライズはグローバルな業界全体でも共有しうる課題であるはずで この分野でデファクトスタンダード確立を目指すことは非常に意義あることだと思います

また日本人/日本語話者/日本文化/日本という地政学的に海に囲まれた地域/国では先に書いたような日本特有の言語的文化的構造が 外からは孤立的傾向を また内にはある種の支配的閉塞的傾向を保持していて 人間が本来持っているはずの自由闊達な活動意欲をスポイルしているという状況があります

この状況を打破しようとするとき 人間だけでなくテクノロジーや概念など関係する対象要素を全てアクターとして同等に扱うアクターネットワーク理論の視点を利用することが 非常に有効に働くかも知れません

さらに長期的には 日本語ローカライズモデルを確立すれば これを敷衍することでより進化したグローバルな包括的汎用大規模言語モデルの再構築につなげることも可能になります

Welcome to Japanese Language World!