

ニュースリリース

## オルツ、LLMハルシネーションの自動評価エンジンを開発 ～ハルシネーション自動スコアリングで、生成AIの誤出力の発生を発見可能に～

P.A.I.®（パーソナル人工知能）をはじめ、AIクローン技術で作り出すパーソナルAIの開発および実用化を行う株式会社オルツ（本社：東京都港区、代表取締役：米倉 千貴、以下、オルツ）は、2024年5月9日に、大規模言語モデル（LLM）におけるハルシネーションをスコアリングする手法の開発に成功したことを発表いたします。



「ハルシネーション」とは、LLMがまるで幻覚を見ているかのように、正当性がなく、事実に基づかない虚偽の回答をしてしまう現象です。このようにLLMが誤った情報を出力することは、企業や個人にとって深刻な信頼問題を引き起こす可能性があるだけでなく、LLMの今後の応用にも大きな障壁となっています。

オルツでは、以前から独自の大規模言語モデルを開発・提供しており、この開発の経験を活かし、ハルシネーション問題を解決する研究開発を続けてまいりました。この度、ハルシネーションを解決する方法の一つとして、ハルシネーションが発生するか否かを判定し、その発生確率（ハルシネーション・スコア）を自動的に評価できる手法を独自に発明し、「ハルシネーション自動評価エンジン」を開発いたしました。

本開発における実験では、本「ハルシネーション自動評価エンジン」は、JcommonsenseQAデータセットから作成した擬似評価セットにおいて72%の正解率でハルシネーションの判定を実現しました。本エンジンは、オルツが独自に開発してきた軽量型大規模言語モデル「[LHTM-OPT](#)」だけでなく、GPT-3.5、Llama2など、様々なLLMのハルシネーションのスコアリングが可能になります。

また、本エンジンは、一貫性を重視してハルシネーションの評価を行います。具体的には、同じ入力データに基づいて複数の生成プロセスを行い、これらの結果を比較します。このアプローチにより、生成された内容における

不一致や矛盾を特定し、それに基づいて「ハルシネーション」、すなわち、学習データや事実に基づかない不正確な生成物が生じているかどうかを確率的に評価します。

「ハルシネーション自動評価エンジン」は、当社の「[alt developer](#)」のAPIサービスを通じて提供を開始いたしました。（alt Developer：<https://developer.alt.ai/api-doc.html#tag/LHTM-OPT/operation/lhtml-opt-completion>）

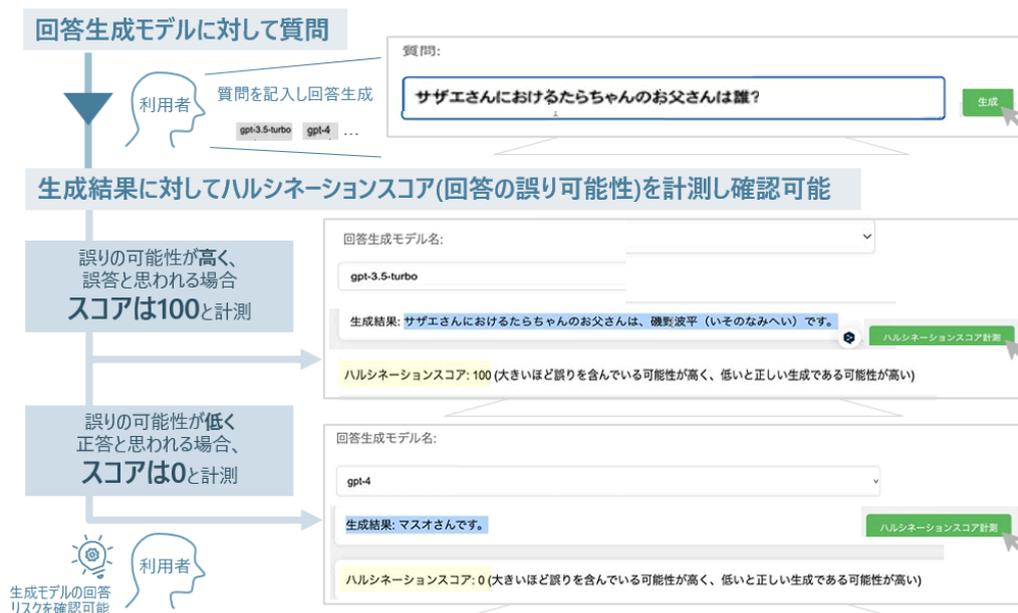
本件ならびにLLMを活用したプロジェクトについては、以下のアライアンス窓口にお問い合わせください。

> ハルシネーションスコア計測アプリケーション デモ動画 <

## ハルシネーションスコア計測アプリケーション

alt+

大規模言語モデルの評価として、質問に対する「回答の不確実性」をスコアリング (API提供も可能)



(<https://youtu.be/-k-SDIPje4>)

▶ **LHTM-2/LHTM-OPT/GPT など大規模言語処理ソリューションに関するお問い合わせ先**

<https://alt.ai/aiprojects/gpt/>

### ■株式会社オルツについて

2014年11月に設立されたオルツは、P.A.I.® (パーソナル人工知能)、AIクローンをつくり出すことによって「人の非生産的労働からの解放を目指す」ベンチャー企業です。生成AI、独自開発LLM及び音声認識技術をはじめとするAI要素技術を豊富に保有し、それらを活用した多くのAI Productsを開発・提供しています。2024年4月までの累計調達額は約100億円超に達しています。

<https://alt.ai/>

<報道関係者からのお問い合わせ先>

株式会社オルツ 広報 西澤

e-mail：[press@alt.ai](mailto:press@alt.ai)

**<アライアンスに関するお問い合わせ先>**

株式会社オルツでは、IT・金融・建設・物流・メディア・製造・小売・サービス業など、ジャンルを問わずAIソリューションの提供および支援を行っております。

お気軽にお問い合わせください。

株式会社オルツ AI Solutions事業部 浅井

e-mail : [gptsolutions@alt.ai](mailto:gptsolutions@alt.ai)