

## ハルシネーションを大幅抑止し専門的な質問にも正確な回答が可能な生成 AI ストックマーク 1,000 億パラメータ規模の独自 LLM を公開

### 商用利用可能な日本最大級の大規模言語モデルを公開

ストックマーク株式会社(本社：東京都港区、社長：林 達、以下：当社)は、2024年2月に、経済産業省と国立研究開発法人新エネルギー・産業技術総合開発機構(以下「NEDO」)が主催する、国内の生成 AI 開発力強化を目的としたプロジェクト「GENIAC」への採択を受け、ハルシネーション<sup>※1</sup>を大幅抑止し、厳密さが要求されるビジネスシーンでも信頼可能な 1,000 億パラメータ規模の大規模言語モデル(Large Language Model、以下：LLM)「Stockmark-LLM-100b」の開発に成功し、本日より各種サイトにて公開しました。

フルスクラッチ<sup>※2</sup>で開発された当該基盤モデルは 1,000 億パラメータ規模であり、国内最大級の基盤モデルとなります。なお、当モデルの開発は、国立研究開発法人産業技術総合研究所との共同研究の一環として行われました。

※1：人工知能が学習したデータからは正当化できないはずの回答を堂々とする現象

※2：システムを作る際に、既存のモデルを用いずにゼロから組み上げる開発手法

#### 【Stockmark-LLM-100b モデル公開先】

- 事前学習モデル：<https://huggingface.co/stockmark/stockmark-100b>
- 指示学習モデル：<https://huggingface.co/stockmark/stockmark-100b-instruct-v0.1>

#### ◆「Stockmark-LLM-100b」について

##### ハルシネーションの抑止例

Q：ストックマークが製造しているエポキシ樹脂の用途をまとめてください<sup>※</sup>

GPT-4-turbo	Stockmark-LLM-100b
ストックマーク (Stockmark) が製造しているエポキシ樹脂は、さまざまな産業や用途で使用されています。一般的な用途としては以下のようなものがあります  1. 接着剤: エポキシ樹脂は強力な接着剤として広く使用されています。金属、プラスチック、ガラス、木材など、様々な素材を接着する際に使用されます。 (以下省略)  これらは一般的な用途であり、エポキシ樹脂はさらに多岐にわたる産業や分野で使用されています。	ストックマークは、エポキシ樹脂の製造をしていないため、用途はありません。

※：当社は自然言語処理を応用し、業界機密の漏洩と意思決定の支援を行うリービンの開発・運営を行う企業であり、エポキシ樹脂の製造は行っていません。

Stockmark-LLM-100b は、公開されている既存の LLM モデルを用いずに、当社がフルスクラッチで開発した独自の LLM です。当 LLM の特徴は、独自に収集したビジネスドメインの日本語データを中心に事前学習を行うことで、日本語・ビジネスドメイン・最新の時事話題に精通していることに加え、“ハルシネーション”の大幅抑止に成功した、厳密さが求められるビジネス領域でも信頼して活用可能な点です。

2022 年以降、生成 AI に対する注目は日々加熱し、中でも LLM については多くの基盤モデルが開発・公開をされてきましたが、いずれのモデルでもハルシネーションは共通の課題として残され、この課題により、間違いが許容されないビジネス領域における生成 AI 活用が大きく阻害されていました。

当社では、Stockmark-LLM-100b の開発において、ハルシネーションの抑止に注力して開発を行ってきました。ビジネス上の基礎的な質問への回答は勿論、製造業などの専門性の高い知識が求められる領域においても専門性を理解し高い精度で回答が出来ることに加え、答えることが出来ない質問に対しては、無理に回答を生成するのではなく、しっかりと「分からない」と答えることが出来る、厳密さが求められるビジネスシーンにおいても信頼して活用いただけるモデルの開発に成功しました。

なお、Stockmark-LLM-100b は、商用利用可能なモデルとしてオープンソースで公開することで、より多くのユーザーにお試しいただくことできるモデルとなっています。

【Stockmark-LLM 紹介サイト】<https://llm.stockmark.co.jp>

【GENIAC 採択、100b 開発に関するプレスリリース】<https://stockmark.co.jp/news/20240202>

## ◆ 「Stockmark-LLM-100b」と他モデルの比較及び実行例

### 1：独自基準での評価

当社が独自に設定した、ビジネス情報や最新の時事話題に関する質問(Stockmark Business Questions)の正答率も 0.9 と 9 割を超える正答率となっており、GPT-4-turbo と比較しても高精度な生成を実現しています。

・ Stockmark Business Questions : <https://huggingface.co/datasets/stockmark/business-questions>

#### Stockmark Business Questions 質問回答精度

モデル	回答精度
Stockmark / Stockmark-LLM-100b	0.90
GPT-4-turbo	0.64
GPT-3.5-turbo	0.44

#### Q：「タイパ」とは？

GPT-4-turbo	Stockmark-LLM-100b
「タイパ」とは、一般的にはタイプライターやコンピューターのキーボードなどで文字を打つこと、またはその速度や技術を指す言葉です。  広義では、タイピングを行う人々やタイピングスキル、タイピング競技会などの文脈で使用されることがあります。また、特定の地域名や人名にも使用される場合があるため、文脈によって意味が異なることがあります。	タイパとは、時間効率を表す言葉であり、Z世代を中心に浸透しています。以下は、タイパに関する重要なポイントです。  - タイパは、Web会議の利用やインスタント食品の利用など、様々な場面で使われます。 - タイパを意識した製品もあり、例えば、SONICの持つタイマーなどがあります。 - タイパを上げるためには、学習タイマーを使用することが有効です。 - タイパに対する意識は世代間で異なり、Z世代以外の世代でもタイパが重要視されています。

### 2：一般ベンチマークでの評価

LLM のモデル評価で活用される VicunaQA Benchmark<sup>※3</sup>においても、他モデルと比較し高い水準を達成しています。

#### Vicuna Benchmarkの評価結果

モデル	Score
Stockmark-LLM-100b	5.97
Llama2 (tokyotech-llm--swallow-70b-instruct-hf)	5.59
GPT-3.5 (text-davinci-003)	5.08

※3：決まった回答が存在しない非定型的なタスクにおける LLM の性能評価を目的としたベンチマークです。  
計算やコーディングなどを要するカテゴリーを除く 60 問で評価  
参照：<https://github.com/ku-nlp/ja-vicuna-qa-benchmark>

### 3：推論速度計測<sup>※4</sup>

各モデルの推論処理速度を計測するために、100 文字の日本語作文を書くのに要した時間を計測したところ、1.86 秒と他モデルと比較しても高い速度性能を有しています。

#### 各モデルの推論速度計測

モデル	推論速度(sec per 100 chars)
Stockmark-LLM-100b	1.86
GPT-3.5-turbo	2.15
GPT-4-turbo	5.48
Llama2-70B / Swallow70B	3.14 / 2.22

※4：100 文字の日本語作文を書くのに要した時間を計測

## ◆「Stockmark-LLM-100b」今後の展開について

今後は、自社プロダクトである市場及び技術情報のプラットフォーム「Anews」に、Stockmark-LLM-100bを実装することで、Microsoft Bing や Google Bard に代わる、新しいビジネス情報の検索体験を提供していきます。さらに、当社生成 AI 基盤の高いセキュリティ性を活かし、社内情報を含めたあらゆるビジネス情報の RAG プラットフォームとして「Anews」を進化させていきます。

また、自社プロダクトへの活用以外にも、当社の生成 AI 基盤を活用した「①RAG に必要な社内外データの構造化及びデータベース構築 / ②より専門的な知識が求められるケースでの企業毎の独自 LLM 開発 / ③LLM を活用したアプリケーション開発」等も提供していきます。

## ◆取締役 CTO 有馬 幸介のコメント



1,000 億パラメータ規模の LLM をゼロから構築し公開が行われた前例は、日本のみならずグローバルでも殆ど存在しておりませんでした。そのため、いざ開発をしてみると 100 億パラメータ規模の頃では起きなかった想定外の事象が多発し、技術的挑戦が日夜続きましたが、限られた時間の中、チームで知恵を絞り、GENIAC 事務局の方々の支援も受けて学習の完了に至りました。

当該モデルやノウハウの公開を行うことで、社会の前進に貢献できれば幸いと感じるとともに、当該モデルの社会実装を進め、人類の前進により一層貢献できるように引き続き頑張りたいと強く感じております。

## ◆GENIAC について

「GENIAC」は、経済産業省と NEDO が主催する、日本国内の基盤モデル開発力を底上げし、また企業等の創意工夫を促すためのプロジェクトです。

計算資源の提供、利活用企業やデータホルダーとのマッチング支援、グローバルテック企業との連携支援やコミュニティイベントの開催、開発される基盤モデルの性能評価など、生成 AI によって世界の変革がもたらされようとしている中、国内外の関係者の知見を結集し、日本の開発力向上を目指します。



## ◆ストックマーク株式会社について

ストックマーク株式会社は「価値創造の仕組みを再発明し、人類を前進させる」をミッションに掲げ、最先端の自然言語処理技術を活用し、多くの企業の事業創造活動を支援しています。

当社は自然言語処理技術を活用し、国内外の約 35,000 サイトのビジネスニュースを分類しています。業務に直結する情報を AI が届けることにより、組織の情報感度を高め、事業アイデアの着想と組織内での発展を促す「Anews(エーニュース)」、AI が組織や業務に合わせて必要な情報を構造化し、示唆が得られる市場調査サービス「Astrategy(エーストラテジー)」を通じて企業の新たな価値創造をサポートします。

会社名 : ストックマーク株式会社

所在地 : 東京都港区南青山一丁目 12 番 3 号 LIFORK MINAMI AOYAMA S209

設立 : 2016 年 11 月 15 日

代表者 : 代表取締役 CEO 林 達

事業内容 : 自然言語処理を活用した、事業機会の探索と意思決定の支援を行うサービスの開発・運営

URL : <https://stockmark.co.jp/>