



カラクリ株式会社



カラクリ、AWS Trainium を活用して日本語処理に特化した大規模言語モデルを開発。GPU よりコスト効率に優れた手法で国産 LLM として最高性能の評価を獲得

導入効果

- 50% : GPU と比べた開発コスト削減率
- 2.5 か月 : 初期開発完了までの期間
- AWS Trainium の活用ノウハウを身に付け、競争力を強化
- 経済産業省・NEDO による国内生成 AI 開発力強化プロジェクトに採択
- 業界認知度の飛躍的向上、技術メディアでの高い注目度獲得

ご利用中の主なサービス

- AWS Trainium
- AWS Inferentia
- AWS ParallelCluster

企業名：カラクリ株式会社

業種：IT サービス

国名：日本

従業員数：78 人 (2024 年 11 月 現在)

ウェブサイト：

<https://about.karakuri.ai/>

概要：カスタマーサポート向けの AI ソリューションを提供するカラクリ。同社は AWS Trainium を活用して、日本語処理に特化した大規模言語モデル (LLM) を開発しました。従来の GPU 環境と比べて大幅に安価なコストで開発された「KARAKURI LM」は、ベンチマークテストにおいて国産 LLM として最高性能の評価を獲得。この成果を受けて、現在は AWS とともに生成 AI の実用化を目指す企業の支援にも取り組んでいます。

ビジネスの課題：LLM 開発における計算リソースとコストの壁

AI チャットボットサービス「KARAKURI chatbot」をはじめ、カスタマーサポートやコンタクトセンター向けの多彩な AI ソリューションを提供するカラクリ株式会社。AI の高度な知見と独自の特許技術を活かし、高精度な AI ソリューションを開発するノウハウは高く評価されています。

近年、話題となっている生成 AI や LLM は、カスタマーサポートの分野においても大きな期待を集めるようになってきました。こうした背景から、カラクリでは「人と AI の共同作業」のさらなる深化を目指し、2023 年 4 月に「カラクリ LLM ラボ」を設立。LLM の自社開発を視野に入れた研究開発を進めてきました。しかし、ここで課題となっていたのが高性能な計算リソースを維持するための高額なコストでした。

取締役 CPO で LLM 開発プロジェクトを率いる中山智文氏は、次のように振り返ります。

「LLM の利用者は最先端の体験を期待しますが、それに応えるためには高額な開発コストが必要です。また、AI の開発競争が激化する中で GPU の供給不足が続いており、その確保も難しい状況にありました」

この状況を一転させたのが、アマゾン ウェブ サービス (AWS) から提案された「LLM 開発支援プログラム」への参加でした。このプログラムでは、計算リソースの確保や事前学習用のクレジットを含む包括的な支援

が提供されます。Research and Development チームの吉田雄紀氏は、次のように話します。

「このプログラムと出会う少し前、社内の勉強会では『LLM の独自開発は理論上可能でも、コスト面での障壁が極めて高い』という結論に達していただけに、このプログラムは当社にとってまさに救世主のような存在でした」

ソリューション：AWS Trainium などを利用して LLM 開発に着手

LLM 開発支援プログラムでは、AWS が開発した機械学習チップである AWS Trainium を利用することになりました。同社にとって馴染みのない AWS Trainium の利用は学習が思うように進まないなどのリスクも想定されましたが、あえて AWS の LLM 開発支援プログラムへの参加を決めた理由について、中山氏は次のように説明します。

「AWS Trainium は GPU よりも供給が安定していることに加えて、AWS の LLM 開発支援プログラムは技術面でのメンタリングのみならず、投資家や社内向けに独自の LLM 開発の意義を説明する際のアドバイスなど、包括的に支援してくれました。丁寧な対話を通じて当社のニーズを支援プログラムに反映していただき、親身に相談に乗ってくださる姿勢も心強かったです」

こうしてカラクリは、2023 年 10 月に日本語に特化したカスタマーサポート向け LLM の実用化プロジェクトに着手。これま





中山 智文 氏
取締役 CPO
LLM 開発プロジェクト
リーダー



吉田 雄紀 氏
Research and
Development



中嶋 恭久 氏
Research and
Development

企業概要

カラクリ株式会社

「今までにないカラクリで世の中を豊かに」をミッションに掲げ、AI アルゴリズムの研究開発を基軸に多彩な AI ソリューションを提供。主力の SaaS 事業ではコンタクトセンターやカスタマーサポート向けの AI チャットボット「KARAKURI chatbot」や FAQ を提供。生成 AI 開発コンサルティング事業では大手企業や研究機関にむけて生成 AI のコンサルティングや個別開発などを手がけている。



「AWS Trainium を活用することで、開発コストを大幅に削減することができました。最初のモデルの開発コストは約 800 万円で、これは GPU を利用した場合の半額程度です」

中山 智文 氏

取締役 CPO LLM 開発プロジェクトリーダー

で培ってきたチャットボット運用のノウハウを活かして高品質な指示と応答データを作成し、個人情報を含まない許諾済みの自社保有データを活用しながら、実用的かつ信頼性に優れたデータセットを構築しました。これらのデータを活用してオープンソースの LLM をベースにチューニングを実施しました。

LLM の学習には機械学習向けの AWS Trainium、推論には AWS Inferentia を活用し、クラスター管理には AWS ParallelCluster を使用。当時 AWS が提供する SDK である AWS Neuron では対応していなかったモデルもあり、カラクリ側独自でモデル対応を進める場面もありました。開発を担当した Research and Development チームの中嶋恭久氏は、次のように話します。

「新しい機構を取り入れた今回のモデルでは、独自のコンパイラプログラムを組み込む必要がありました。当初は対応が難しい部分もありましたが、AWS のスペシャリストとの密にコミュニケーションを通じて解決することができました」

導入効果:国産として最高性能の LLM 開発に成功、生成 AI の普及推進へ

2024 年 1 月、カラクリは国内最大規模となる 700 億パラメーターの LLM「KARAKURI LM 70B v0.1」を発表しました。同時に LLM のベンチマークテスト「Japanese MT-Bench」で性能評価を実施し、モデル公開時点で国産 LLM としては最高性能を達成しました。そして、同年 5 月には第 2 弾を、6 月には第 3 弾の AWS Trainium を利用して学習したモデルを発表しました。

AWS Trainium によって開発コストを想定より大幅に削減できた点について、中山氏は次のように話します。

「最初のモデルの開発コストは約 800 万円で、これは GPU を利用した場合の半額程度です。その後のモデルは 30 万円、3 つめの AI エージェントモデルは 75 万円で開

発することができました。準備を含めた開発コストの総額は 2,000 万を超えていますが、一部は支援プログラムからのサポートを受けられたので大変助かりました」同社の高性能 LLM の開発はメディアでも取り上げられ、知名度も向上。2024 年 10 月には、経済産業省と NEDO が協力する国内生成 AI 開発力強化プロジェクト「GENIAC (Generative AI Accelerator Challenge)」にも採択されました。同社が取り組んだのはカスタマーサポート向けのモデル開発ですが、特に需要が集中する GPU に比べて計算リソースを安定的に確保できる AWS Trainium を活用した同社の開発手法には大きな注目が集まっています。

「今回のプロジェクトで質の高いモデルを開発できたことには一定の手応えを感じています。特に、どのようなデータを学習させれば効果的か、汎用性を保ちながら特定のスキルを向上させる方法など、実践的なノウハウも蓄積することができました」(中嶋氏) さらに同社では現在、これらのノウハウをもとにテキストだけでなく、画像認識、音声認識、音声発話といった多様な要素を統合的に扱う「オムニモーダル」モデルの開発にもチャレンジしています。

「昨今の LLM 開発においては、画像や動画の作成を効率化する仕組みやクラウドインフラの知識不足が障壁になるケースがありますが、これらのハードルをクリアすることができれば、より多くの企業や研究者が LLM 開発に参入できるようになります。この基盤整備において、AWS の今後の取り組みには大きな期待を寄せています」

カラクリは現在、AWS が提供する「生成 AI 実用化推進プログラム」に協業パートナーとしても参画。生成 AI の実用化ビジネスにおける次のステップへと踏み出しています。「AWS と協業することで、これまで手が届かなかった領域にまでビジネスを拡大することができます。今後も AWS とのパートナーシップを強化しながら、新たなビジネス課題にチャレンジしていきたいと思えます」(中山氏)



アマゾン ウェブ サービス ジャパン合同会社

〒141-0021 東京都品川区上大崎 3-1-1 目黒セントラルスクエア <https://aws.amazon.com/jp/>

Copyright © 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.