

News Release



報道関係各位

2023年8月23日

東京エレクトロン デバイス株式会社

大規模言語モデルの学習をクラウド上で利用可能

Cerebras「AI Model Studio サービス」をハイレゾ社と共同で提供開始 生成 AI の開発を促進

東京エレクトロン デバイス株式会社（横浜市神奈川区、代表取締役社長：徳重 敦之 以下、TED）と株式会社ハイレゾ（本社：東京都新宿区、代表取締役：志倉 喜幸）は AI 学習の課題解決を目指して協業し、大規模言語モデルの学習をクラウド上で利用可能な Cerebras Systems（セレブラスシステムズ）「AI Model Studio サービス」を 2023 年 8 月 23 日より提供開始します。

生成 AI の研究や活用が進む中、OpenAI の GPT3/GPT4 などディープラーニング技術を用いて構築されたさまざまな大規模言語モデルが登場しています。日本国内でも ChatGPT などの生成 AI サービス活用による業務効率化、新規事業開発、DX 推進、社員教育などさまざまな効果が期待されています。

一方で、ChatGPT などの利用には回答内容の信頼性、機密情報の漏えいや著作権侵害、個人情報の流出などの懸念があり、ビジネスでの利用は制限されているのが現状です。

ビジネスで利用できる安全な生成 AI を独自に開発するには、占有環境で大規模言語モデルをファインチューニングし、自社データを用いて学習させる必要があります。しかしながら、大規模言語モデルの学習には GPU など高速な計算処理能力を持つハードウェアリソースを多数用意しなければならず、インフラ資源の確保が課題となっています。

■ 「AI Model Studio サービス」の概要

「AI Model Studio サービス」は、大規模言語モデルの学習環境をクラウドで提供することで、生成 AI の開発を促進するサービスです。クラウドを通じ、セキュアな環境でウェハースケールエンジン (WSE-2) を搭載した超高速ディープラーニングシステム「CS-2」を占有的に利用できます。

本サービスは、13 億パラメータから最大 1750 億パラメータまでの最先端の GPT クラスのモデルを選択して利用することができます。ユーザーが同等の環境を一般的なパブリッククラウドやハードウェアリソースを用意した場合と比較して、利用費用を約半額に抑え、大規模言語モデルの学習コストの削減に寄与します。

また、データセットを用意するだけで、GPT クラスのモデルや Cerebras Systems「Cerebras-GPT」な

どのオープンソース大規模言語モデルを、数行のコマンドでトレーニングすることができます。CS-2 クラスターの処理速度を生かし、実行処理時間の短縮が可能となり、生成 AI 開発の迅速化に貢献します。

(参考) Cerebras AI Model Studio URL : <https://www.cerebras.net/product-cloud/>

■ 「AI Model Studio サービス」の主な特長

- ・セキュアな環境で自社データ学習を実行可能
- ・同規模環境をパブリッククラウドで構築する場合と比較し、利用費用は約半額・8 倍の学習パフォーマンスを実現
- ・数行のコマンドで GPT クラスのモデルをトレーニングできる環境を提供
- ・13 億パラメータから最大 1750 億パラメータまでの最先端の GPT クラスモデルを選択可能
- ・Cerebras と連携し、TED エンジニアが「CS-2」の利用を技術サポート
- ・最小限のデータセットの用意で「AI Model Studio サービス」の利用が可能なファインチューニングオプションを提供

■ 提供価格

学習モデルに基づき提供します。

Model	Parameters (B)	Tokens to train to Chinchilla point (B)	Cerebras Model Studio days to train	ご提供価格
GPT-3 XL	1.3	26	0.4	¥412,000
GPT-J	6	120	8	¥7,425,000
GPT-3 6.7B	6.7	134	11	¥6,600,000

TED は、これまで培ったネットワーク、ストレージシステムと、AI やクラウドなど最新鋭の製品・サービスと高度な技術サポートの提供により、お客様ビジネスの拡大に貢献してまいります。

● 「Cerebras-GPT」について

AI アプリケーションのプラットフォーム Hugging Face、ソフトウェア開発のプラットフォーム GitHub で Apache 2.0 ライセンスの下で公開されているオープンソース GPT モデルです。EleutherAI の Pythia を補完するように Cerebras 独自に設計されており、1 億 1,100 万から 130 億のパラメータまで、7 種の学習済み大規模言語モデルを無償で利用できます。また、こちらは商用利用も可能となっております。

東京エレクトロン デバイス株式会社について

東京エレクトロンデバイスは、半導体製品や IT ソリューション等を提供する「商社ビジネス」と、お客様の設計受託や自社ブランド商品の開発を行う「メーカー機能」を有する技術商社です。

URL: <https://www.teldevice.co.jp/>

CN BU では、ストレージ、ネットワーク、セキュリティ製品、および AI プラットフォームまで幅広く取り扱い、クラウド時代のビジネスソリューションとしてご提案しています。世界の最先端テクノロジーをいち早くキャッチするマーケティング機能を持ち、お客様に最適なソリューションを導入からサポートまで一貫して提供しています。

CN BU サイト：<https://cn.teldevice.co.jp/>

<本件に関する報道関係からのお問合せ先>

東京エレクトロン デバイス株式会社 総務部広報グループ

お問い合わせフォーム：<https://www.teldevice.co.jp/cgi-bin/form/contact.php>

<本製品に関するお客さまからのお問い合わせ先>

東京エレクトロン デバイス株式会社

CN BU CN ビジネス開発室

お問い合わせフォーム：<https://share.hsforms.com/1G6U-ToHuTfubPI3xlxWBcQbz1ou>

※ このニュース リリースに記載されている会社名、製品名は、各社の登録商標または商標です。