

導入事例

インテル® Gaudi® 3 AI アクセラレーター

intel

選択肢が広がるインテル® Gaudi® 3 AI アクセラレーターの採用

neoAI Chat が描く近未来の生成 AI 基盤



株式会社 neoAI

所在地：東京都千代田区神田須田町 2-5
東京神田須田町ビル 2 階

設立：2022 年 8 月

資本金：約 5,600 万円（資本準備金を含む）

事業内容：法人向け AI ソリューション事業、自社 AI サービス事業

<https://neoai.jp/>

生成 AI の急速な発展により、企業における AI 活用の可能性も広がり続けています。日々新しい AI サービスが展開されている一方で、いかなる AI をインテグレーションすべきか頭を悩ませている企業も少なくありません。とりわけ日本企業においてはドキュメントを LLM に学習させるには多くの問題があります。

こうした課題に対し、生成 AI オールインワン・パッケージ「neoAI Chat」を展開しているのが neoAI です。同社の代表取締役 CEO & Founder である千葉 駿介 氏に、neoAI Chat の特徴や、このたび採用を決定したインテル® Gaudi® 3 AI アクセラレーターについて伺いました。



株式会社 neoAI 代表取締役 CEO 千葉 駿介 氏

AI に対しドキュメントを“翻訳”する neoAI Chat

ChatGPT 登場直前の 2022 年 8 月、AI やディープラーニングを研究する「東京大学 松尾研究室」発のスタートアップ企業として設立された neoAI。千葉氏は立ち上げメンバーである東京大学の学部生 6 人の代表として「neoAI Chat」に関わってきました。

同氏は今年 5 月に発表された「Forbes 30 UNDER 30 2025 Asia」において、neoAI で COO を務める寺澤 滉士良 氏とともに“世界を変える 30 歳未満”30 人に選出されるなど、いまテック業界で注目されている人物のひとりです。





neoAI Chat

neoAIは、生成AIの技術をビジネスの現場で適時に応用することを目指しており、さまざまなアプリケーション、ソフトウェアの提供とインテグレーション・サービスを行っています。

「最近ではAIエージェントなども話題になっていますが、AIの技術がどんどん複雑になってきているなかで日本の企業が抱えているのは、日々更新されていく技術を、セキュリティを担保しながら、どう社内にインテグレーションするかです。そこに対するサービスとして、私たちはneoAI Chatを提供しています」(千葉氏)

同社のneoAI Chatを導入しているのは、銀行や信用金庫、電力会社といった高いセキュリティが求められる企業がほとんどです。特に近年、企業からは“検索拡張生成(RAG)”の需要が高まりつつあります。RAGとは、企業が保有するさまざまなドキュメント・データを大規模言語モデル(LLM)に繋げていく技術であり、いわば企業の社内知識を持ったAIをつくる取り組みと言えます。

「ただ、“言うは易し”で、実際にやってみるとなかなか高い精度が出なかったり、AIが嘘をついてしまったりすることがあります。特に日本企業が持っているドキュメントは日本語ベースであるだけでなく、フローチャートのような図が入っていたり、表計算ソフトのセルが結合されていたり、手書きのものもあつたりします。また業界によっては専門用語も使用されており、汎用的な製品ではこれらを一気に取り込むのは難しいのです。そういったさまざまなドキュメントをAIが読めるように前処理を行うのが、我々のneoAI Chatです」(千葉氏)

RAG開発は非常に手間がかかる作業ですが、neoAI Chatならば企業がドキュメントをアップロードするだけで前処理のアルゴリズムが走り、適切なデータがAIに入力されます。いわばneoAI Chatは、AIに対して企業のドキュメントを翻訳してくれる“通訳”のような存在です。

クラウド型サービスのみならず 多種多様な生成AI基盤で利用が可能

AIを管理する機能も搭載されています。現在、すでにOpenAIやGoogle、Metaなどが開発したLLMが存在しており、それぞれが得意

分野を持っています。今後はさまざまなAIが乱立する世界になっていくことは疑うべくもないでしょう。

しかし、AIは一般的に使用ごとにコストがかかり、企業にはこれから、コストや機能を踏まえたLLMのマネジメントが求められることになると予測されます。neoAI Chatはそういった次の世界に先駆け、さまざまなLLMを一括で使えるようにし、ユースケースごとに繋げたり付け替えたりできる機能を有しています。

「我々がやっていることは、実はすごく泥臭いものです。単にLLMを選ぶだけではなく、お客様特有のドキュメントが読めるように個別に作り込んでいます。結果として、その企業に特化したようなものができあがります。とはいえneoAI ChatはSaaSなので、その業界全体で使えるように標準化を行ったうえで実装しています」(千葉氏)



AIを管理し、ユースケースごとに利用できることはセキュリティの向上にも役立ちます。例えば、個人情報保護法や企業のプライバシー/セキュリティ・ポリシーによって、クラウドサービスにアップロードできないデータは少なくありません。

見ればAIはお金がかかる投資です。コストと性能のバランスが良ければ企業も採用しやすく、また neoAI としても勧めやすくなります。LLM だけでなくチップも選べる時代に向け、インテル® Gaudi® 3 AI アクセラレーターは十分な価値を有していると千葉氏は評価しています。

インテル® Gaudi® 3 AI アクセラレーターは有効な選択肢の1つに

neoAI Chat はもともと SaaS からスタートしたプロダクトですが、シングルテナントやオンプレミスなどへの対応を進めながら順調にバージョンアップを進めています。インテル® Gaudi® 3 AI アクセラレーターでの技術検証を終えた今、千葉氏に今後の展望を伺いました。

「次のステップとして、インテルやハードウェア・ベンダーと話し合いながら、お客様にどのようなパッケージを提案していくかを固めていくことになるでしょう。我々がすぐにインテル® Gaudi® 3 AI アクセラレーターの検証を行えたように、検証自体はすぐに始められる環境があります。なかでもコスト面が課題になっている企業にとっては、インテル® Gaudi® 3 AI アクセラレーターは有効な選択肢になるのではないのでしょうか」(千葉氏)

最後に千葉氏は、いち研究者、そして、いちエンジニアの視点から、今後について以下のように展望を語ってくれました。



「用途ごとにチップを選べるということは、いち技術者としてすごく嬉しいなと思っています。LLM も最初は OpenAI が出てきて、それが今では多数の LLM から選ぶことができます。今後は選択肢がより増加し、コストと性能とスピードのバランスを取りながら最適なものを選ぶようになるでしょう。それらを吟味し、エンドユーザーの最適を提案する立場にある私たちにとって、インテル® Gaudi® 3 AI アクセラレーターは頼もしい存在となってくれればと期待しています」(千葉氏)

neoAI
<https://neoai.jp/>

neoAI Chat
<https://neoai.jp/neoaichat>

インテル® Gaudi® 3 AI アクセラレーター
<https://www.intel.co.jp/content/www/jp/ja/products/details/processors/ai-accelerators/gaudi.html>



性能は、使用状況、構成、その他の要因によって異なります。詳細については、<https://www.Intel.com/PerformanceIndex/>(英語)を参照してください。

性能の測定結果は、構成に示されている日付時点のテストに基づいています。また、現在公開中のすべてのアップデートが適用されているとは限りません。構成の詳細については、補足資料を参照してください。絶対的なセキュリティを提供できる製品やコンポーネントはありません。

実際のコストや結果は異なる場合があります。

インテルのテクノロジーを使用するには、対応したハードウェア、ソフトウェア、またはサービスの有効化が必要となる場合があります。

インテルは、サードパーティーのデータについて管理や監査を行っていません。ほかの情報も参考にしてデータの正確さを評価してください。

Intel、インテル、Intel ロゴ、その他のインテルの名称やロゴは、Intel Corporation またはその子会社の商標です。

その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

テックプラス (2025 年 6 月 30 日) に掲載されたコンテンツから抜粋し、再構成したものです。

インテル株式会社

〒100-0005 東京都千代田区丸の内 1-4-1 丸の内永楽ビル 25 階

<http://www.intel.co.jp/>

©2025 Intel Corporation. 無断での引用、転載を禁じます。

2025 年 7 月