NVIDIA Elite Partner　GDEP Advance

暗黙知まで理解するLLMの実現をNVIDIA AI Enterpriseが後押しするストックマーク様

自社でLLMや生成AIの技術を開発し、国家プロジェクトにも参画するストックマーク。
壮大な夢の追求と求めるクオリティの実現には「NVIDIA AI Enterprise」が欠かせない。

飛躍的な進歩にともない、さまざまなシーンで活用されるようになった「生成AI」。世界各国がその開発にしのぎを削るなか、日本では経済産業省とNEDO(新エネルギー・産業技術総合開発機構)が国内生成AIの持続的な開発力の向上や社会実装の加速を目的とするプロジェクト「GENIAC」(Generative AI Accelerator Challenge)を始動。生成AIの開発力強化に必要な計算資源の調達やデータセットの蓄積、ナレッジの共有などを支援する事業を行っている。

このGENIACの支援事業において、支援対象となる採択事業者に2024年2月の第1期から連続で選ばれているのが、国内AIのリーディングカンパニーであるストックマークである。

「価値創造の仕組みを再発明する」をミッションに掲げるストックマークは、独自の自然言語処理(NLP)や大規模言語モデル(LLM)、生成AIなどの技術開発に取り組むスタートアップだ。あらゆる業務のAI化を専門チームが伴走支援する「SAT(Stockmark A Technology)」と研究・開発現場の意思決定を支援する製造業R&D向けAIエージェントの「Aconnect」をメイン事業として展開し、企業の意思決定や事業創出などを支援している。


ストックマーク 取締役CTO 有馬幸介氏

ストックマークの創業者で取締役CTOを務める有馬幸介氏によれば、例えば製造業の研究・開発を担う技術者は「業務時間の約35%しか研究・開発に費やせず、残りの約65%は情報収集に使っている」そうだ。このような背景から、同社はSATやAconnectによって情報収集等の業務効率化を実現し、技術者が本来すべき業務である研究・開発により多くの時間を費やせるようサポートしている。

ハルシネーション抑止から暗黙知の読解も可能なLLM

一方GENIACでは、まず第1期で「1,000億パラメータ規模の独自LLM開発」に取り組んだ。有馬氏は「この開発では、公開済みのLLMを使用しなかった点が重要なポイント。フルスクラッチで事前学習を行い、ビジネス領域に強くハルシネーションも大幅に抑止(抑止率90%以上)した独自LLMの開発を目指した」と説明する。そしてその成果は、1,000億パラメータのLLM「Stockmark-LLM-100b」として2024年5月に一般公開された。

さらに第2期では、「ハルシネーションを抑止したドキュメント読解基盤モデルの構築」をテーマに、企業の業務で頻繁に扱われる複雑なドキュメント(テキストだけでなく画像や図版なども含むもの)を正確に理解できるマルチモーダルの基盤モデル開発に挑戦。Stockmark-LLM-100bの開発を通して得た知見や課題を活かすとともに、図/表/画像などを含む複雑なドキュメントを豊富に学習させることでマルチモーダル化させた基盤モデルを開発し、こちらは2025年6月に「Stockmark-2-VL-100B」として公開された。

GENIACの製造業データ等のAI-Ready化に関する研究開発では、伊藤忠商事や神戸製鋼所、住友化学、帝人、ライオンなど日本を代表する16社と協業
(ストックマークのプレスリリースより)

ただし、Stockmark-2-VL-100Bでも上手く読み解けないドキュメントがあり、その問題を解くカギとして有馬氏が注目したのが「データ化はされていないが、人類の叡智のような情報が詰まっているもの。いわゆる“暗黙知”と呼ばれるような要素」だった。

そこで第3期ではこれを踏まえ、ビジネスシーンの各所に点在する「暗黙知」の抽出・形式知化を実現する「“暗黙知”を抽出する業界特化のドキュメント読解基盤モデル」の開発をスタート。さらに2026年5月に採択された「製造業データ等のAI-Ready化に関する研究開発(以下、AI-Ready化事業)」では、日本を牽引する大企業16社と協業し、各企業内に眠る非公開の“秘匿データ”や熟練者のノウハウなどをはじめとする“匠の知”を、AIが学習・活用できる形式に変換

する「AI-Ready化」の実証実験を開始した。これらの暗黙知をデータ化し、より専門的な情報も理解可能なLLMを実現することで「日本企業の生成AI活用力の底上げと、実社会への実装促進に貢献していく」と有馬氏は語る。

このように、自社の事業やGENIACでの取り組みでさまざまな開発を進めているストックマーク。その過程では当然さまざまな課題にぶつかっているわけだが、その課題解決の大きな推進力となっているのが、NVIDIAのエンタープライズ向けAIプラットフォーム「NVIDIA AI Enterprise」である。

効率的な前処理や合成データ生成を実現するNVIDIAプラットフォーム

NVIDIA AI Enterpriseは、AIの開発・運用に必要なツールやライブラリ、サポートなどを提供するプラットフォーム。ストックマークはこのNVIDIA AI Enterpriseで提供される「NVIDIA NIM™」や「NVIDIA NeMo™」などを利用しているが、きっかけはそれまでに抱えていた課題感にあった。

例えば、以前より高性能なAIを開発した場合、そのAIはより多くのデータをさばけるようになる。ただし、膨大なデータをそのままAIに渡してしまうと、AIの応答速度が低下したり、回答精度が低下したりする。そのため、膨大なデータを前処理して有望なデータ順に並び替える「リランキング」が、AIの利活用には「重要なポイントになる」と有馬氏は指摘する。そして、このリランキングを手間なく効率的に実行してくれるのがNVIDIA NIMである。

NVIDIA エンタープライズ事業部 ソフトウェア ビジネスデベロップメントの由良直之氏いわく、NVIDIAでは自社の知見をベースにリランキングモデルを作成しているが、NVIDIA NIMでは「ただ単にそのモデルを提供しているわけではない」と語る。具体的には「NVIDIA NIMにより、GPU最適化済みの推論環境を短期間で構築でき、AI開発期間の短縮につながる」と解説する。

そもそも、オープンソースにも当然リランキングモデルは存在するが、それらはGPUに対して最適化されていない。そのため「高速処理を実現させるためには、時間をかけて最適化する必要がある」と由良氏は補足し、NVIDIA NIMの強みをさらに強調する。

一方でNVIDIA NeMoは、LLMのファインチューニングのスピードアップや学習データの生成などに活用。例えば、図面やフローチャートが含まれる営業資料のようなデータは、インターネット上にたくさん公開されているわけではないことから、そうしたデータを人工的に合成する際に「NVIDIA NeMoを使うことで素早く大量に生成できるようになる。これにより、その領域の学習を加速化できる」と有馬氏は語る。

さらに、NVIDIA NeMoを使った合成データの生成は、GENIACのAI-Ready化事業で進めているAI-Ready化プロジェクトにも利用可能。協業する16社の秘匿データや匠の知も決して豊富ではないことから「本番実装に向けた段階で精度を上げる際には活用できる可能性がある」と期待を寄せる。

そのほか、近年はデータガバナンスの観点から、企業によっては「自社データの利用をオンプレミスのみに限定する」というケースも珍しくない。その点について由良氏は、「NVIDIA AI Enterpriseは、NVIDIAのGPUが使われていればオンプレミスでもクラウドでも同じように動作する」と補足。そのため、ストックマークがNVIDIA AI Enterpriseを使って開発したLLMやサービスは「ハイブリッドに顧客へ提供することが可能」とそのメリットを示す。有馬氏も「当社のAconnectはクラウド経由で提供しているが、プライベート環境で使いたいというニーズは少なくない」と言及。そういったニーズにSATで応えている案件も増えているそうで、オンプレミスでの利活用ニーズは今後高まるだろうと予想する。

NVIDIA エンタープライズ事業部
ソフトウェア ビジネスデベロップメント 由良直之氏

互いにWin-Winの関係性を構築し今後も二人三脚でサービス提供を

このように、ストックマークはNVIDIA AI Enterpriseを使ってさまざまな開発に取り組んでいるが、今後のビジョンのセンターピンとして有馬氏が挙げたのは「ハルシネーションの抑止」である。「そこはまだ決着がついていない」と強く語り、「100%を目指してこれからも続けていく」と決意を示す。

また長期的なビジョンとして挙げたのは「創造性の拡張が可能なAI」の実現だ。こちらも実用レベルを目指して「AIが出したアイデアを完全に信用できるように突き詰めていきたい」と力説した。

そしてこれらの実現には、やはりNVIDIAの力が欠かせない。それだけに有馬氏は、どんなに複雑な構造のモデルやAIが開発されても、それをきちんと動作させられるGPUやツールの提供を期待する。由良氏も「当社が目指しているのは“開発者のワークロードを高水準で効率化すること”。こちらとしてもストックマークから迅速にフィードバックをいただけており、Win-Winの関係性を築けていると感じる。今後も二人三脚で、顧客により良いサービスを提供していきたい」と笑みを見せた。

ストックマークの使用ソリューション

NVIDIA AI Enterprise

AI開発向けのマイクロサービス、フレームワーク、ライブラリに加え、高度なGPUオーケストレーション機能とインフラ管理機能を統合した、包括的な商用ソフトウェアスイート。フルサポート付きで、本番環境への導入に対応。主要なオープンソースツールやAIモデルを安心して導入・運用できるほか、生産性向上と価値創出までの時間短縮を実現する。

NVIDIA Elite Partner　GDEP Advance

株式会社ジーデップ・アドバンス
〒104-6205
東京都中央区晴海一丁目8番12号　晴海アイランド トリトンスクエア オフィスタワーZ棟5階
TEL：03-6803-0620　URL：https://www.gdep.co.jp/