Nomura Research Institute Group

NEWS RELEASE

2024 年 1 月 11 日

株式会社野村総合研究所

NRI デジタル株式会社

NRI グループ、データ漏洩リスクを極小化した個別企業向け生成 AI ソリューション「プライベート LLM」を 2024 年に提供予定

～機密・機微情報を安全に扱え、個別企業の業務に合わせてカスタマイズが可能～

株式会社野村総合研究所（本社：東京都千代田区、代表取締役会長 兼 社長：此本 臣吾、以下「NRI」）と NRI デジタル株式会社（本社：神奈川県横浜市、代表取締役社長：雨宮 正和、以下「NRI デジタル」）は、生成 AI の活用にあたって大きな懸念事項である、「データ漏洩リスク」を極小化する生成 AI ソリューション「プライベート LLM（大規模言語モデル）」（以下「本ソリューション」）を、2024 年春以降に提供する予定です。本ソリューションは、金融機関など、特に高いレベルの情報セキュリティ統制を必要とする企業が利用することを想定しています。

主な特長は、以下の通りです。

■ 機密・機微情報を安全に扱えるよう、プライベートクラウドやオンプレミス環境で動作

生成 AI の活用においては、OpenAI 社[注1]の GPT-4[注2]などに代表される外部サービス型 LLM にどこまで機密・機微情報を送信してよいのか、という点が大きな懸念事項となっています。本ソリューションでは、Meta 社[注3]の Llama 2[注4]を始めとする基盤モデルが公開された LLM を、NRI のデータセンターで稼働するプライベートクラウドサービスや企業自身が情報システムを保有し運用するオンプレミス環境で動作させることで、機密・機微情報を安全に扱うことを可能にします。金融機関などが求める、高レベルの情報セキュリティ統制にも対応可能です。

■ 個別企業の業務に合わせて LLM をカスタマイズ

Llama 2 など基盤モデルが公開された LLM の性能は、現時点では外部サービス型 LLM には及ばないものの、企業が持つデータを利用してカスタマイズ（プリトレーニング[注5]、ファインチューニング[注6]など）することで、タスク内容によっては業務に適用可能な水準の性能を発揮することが期待できます（後述の社内実証結果や参考資料を参照）。本ソリューションでは、より機密性が高い学習時点のデータの漏洩リス

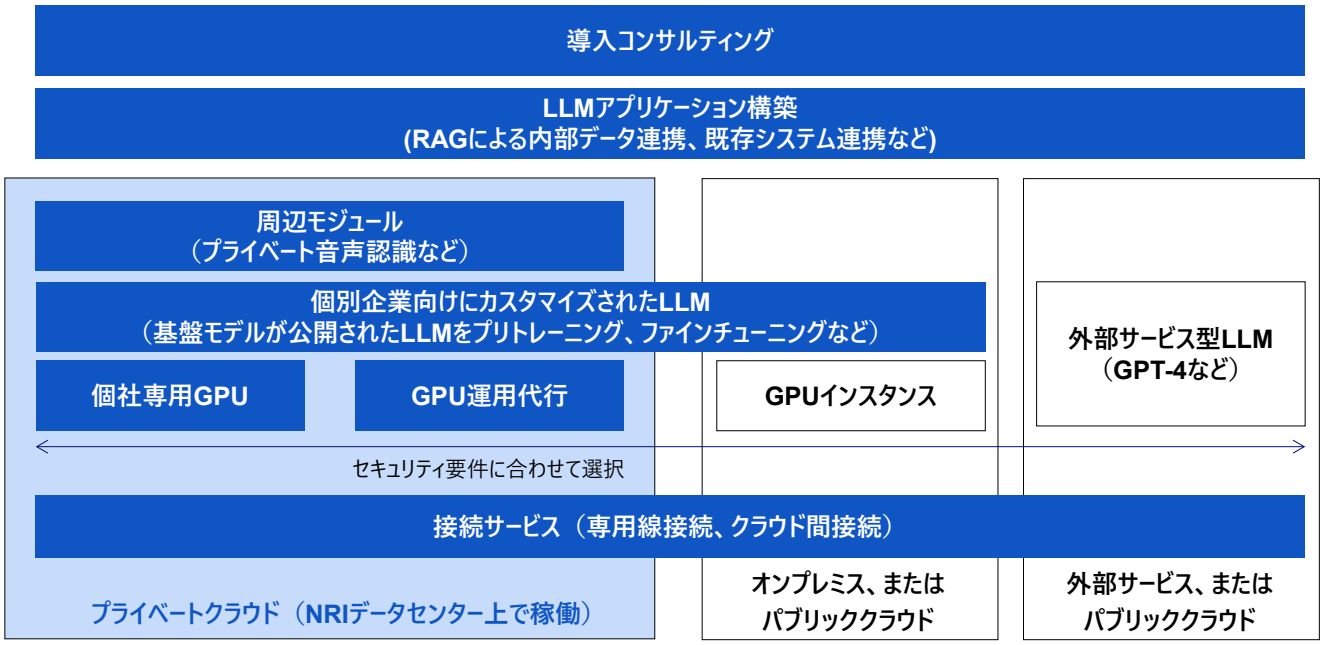
クも極小化した上で、各社の業務に最も適した形で LLM をカスタマイズします。

■ 「プライベート音声認識」などの周辺モジュールも提供

LLM に音声認識機能を組み合わせると、例えばコンタクトセンターや対面での問い合わせ対応など、適用できる業務の幅が大きく広がるため、個人を特定することが可能な音声データの漏洩リスクを極小化する「プライベート音声認識」モジュールなど、周辺モジュールも提供していきます。

本ソリューションのメニュー体系は、下図を参照してください。

図：「プライベート LLM」ソリューションのメニュー体系



白抜き文字部分を NRI が提供。
GPU：Graphics Processing Unit の略称。LLM を動作させるために必要。
RAG：Retrieval Augment Generation の略称。外部の事実データを参照して LLM に正確に回答させる手法。

■ NRI 社内業務を対象として、性能と効果を確認

本ソリューションの性能を検証するため、NRI において、社内の会計事務手続きサポート業務に適用しました。カスタマイズ用の学習データ 6 万件を用いて、NRI のデータセンターに設置した GPU 上で動作する Llama 2 のファインチューニングを行ったところ、業務に適用可能な水準まで性能が向上しました。その結果、本ソリューションによって当該業務の一部を代替し、Q&A 作成の作業時間を 60%削減できました。

NRI と NRI デジタルは、今後も企業や個人のニーズに合わせて最適化された生成 AI を活用できるよう、各種サービスの開発に積極的に取り組んでいきます。

※文中の社名、サービス名、登録商標および商標は、それぞれの会社に帰属します。

注1 OpenAI 社：AI の研究、導入を行っている企業。詳細は次のウェブサイトをご参照ください。
https://openai.com/

注2 GPT-4：OpenAI 社が開発・提供している大規模言語モデルの一つ。

注3 Meta 社：詳細は次のウェブサイトをご参照ください。
https://about.meta.com/

注4 Llama 2：Meta 社が開発した大規模言語モデル。基盤モデルを含めて商用利用可能な形で公開されている。

注5 プリトレーニング：LLM に汎用的な言語パターンや知識を学習させること。

注6 ファインチューニング：LLM を特定のタスクに特化して微調整すること。

【ニュースリリースに関するお問い合わせ】

株式会社野村総合研究所　コーポレート・コミュニケーション部　梅澤、船山

TEL：03-5877-7100　E-mail：kouhou@nri.co.jp

【本件に関するお問い合わせ】

株式会社野村総合研究所　AI ソリューション推進部　岡田

NRI デジタル株式会社　DX 企画　中村

E-mail：private-llm@nri.co.jp

【ご参考】

表 1：外部サービス型 LLM と本ソリューション「プライベート LLM」の特長比較

	外部サービス型LLM	本ソリューション「プライベートLLM」
動作するLLM	OpenAI社GPT-4などに代表される 基盤モデルが公開されていないLLM	Meta社Llama 2などに代表される 基盤モデルが公開されたLLM
提供形態	外部サービス	プライベートクラウド オンプレミス パブリッククラウド
性能	その時点での最高水準	外部サービス型に劣る（徐々に接近）
セキュリティ	パブリッククラウドと同等	オンプレミス、プライベートクラウドと同等 （高レベルのセキュリティ統制に対応可能）
カスタマイズ性	カスタマイズ範囲が制限されている	自由にカスタマイズ可能

表 2：カスタマイズによる Llama 2 の性能向上見通し

タスクの種類	ユースケース	Llama2-7B	Llama2-70B
質問応答（RAGなし）	ヘルプデスク、コンタクトセンター	＊＊	＊＊
質問応答（RAGあり）	ヘルプデスク、コンタクトセンター	＊＊	＊＊＊
要約	応対記録や会議議事録の作成	＊＊	＊＊
文章からQ&A生成	FAQの作成支援	＊＊	＊＊＊
自然文生成	マニュアルや案内文の作成	＊＊	＊＊
対話	アドバイス、カウンセリング	＊＊	＊＊
文体変換	顧客対応支援(丁寧な表現など)	＊＊	＊＊
感情分析	顧客対応やSNS等の感情分析	＊＊＊	＊＊＊
文書分類・タギング	文書の属性分類やタギング	＊＊＊	＊＊＊
異常検知	コールログやメールのコンプライアンス監査	＊＊	＊＊＊
コード生成	アプリ開発、データ分析	＊＊	＊＊＊

7B=70億、70B=700億（パラメータ数）

＊＊＊＝業務適用可能

＊＊　＝課題あり