

## セレンス、言語モデル「CaLLM™」ファミリーの機能向上に向け NVIDIAとの提携を拡大

Janvier 16, 2025

～NVIDIAのオープンソースライブラリ「TensorRT-LLM」と「NVIDIA NeMo」フレームワークを含むソフトウェアプラットフォーム「NVIDIA AI Enterprise」を活用し性能を最適化～

自動車業界向けにAI技術を提供するCerence Inc. (以下セレンス)は、クラウドベースの「CaLLM™(Cerence Automotive Large Language Model)」や組み込み型小規模言語モデル(SLM)である「CaLLM Edge」をはじめとする同社のCaLLM言語モデルファミリーの機能向上を目的として、NVIDIAとの提携を拡大したことを発表しました。これにより、CaLLMは、エンドツーエンドのクラウドネイティブなソフトウェアプラットフォームである「[NVIDIA AI Enterprise](#)」を活用するとともに、CaLLM Edgeの一部の機能は「[NVIDIA DRIVE AGX Orin](#)」によって強化されています。



クラウド型と組み込み型の両方でエージェントフレームワークを車内の会話と統合するには、ハードウェア、ソフトウェア、ユーザーエクスペリエンスの分野の専門知識を組み合わせる、包括的かつ分野横断的な取り組みが必要となります。セレンスは、NVIDIAのハードウェアおよびソフトウェアのエンジニアと協力することで、自動車向け生成AIイノベーションの製品化と生産スケジュール達成の能力強化を図りました。具体的には、[NVIDIAの「TensorRT-LLM」](#)や「[NVIDIA NeMo](#) (生成AIアプリケーションの構築、カスタマイズ、本番運用への展開を可能にするエンドツーエンドのフレームワーク)」を含むソフトウェアプラットフォーム NVIDIA AI Enterpriseを活用することで、CaLLMの開発と導入を迅速化しました。それらを通じて、以下のとおりCaLLMファミリーの最適化とカスタマイズを実現しました。

- NVIDIAにより高速化したコンピューティングとSoCで、車載アシスタント性能の高速化を実現
- [NVIDIAの「NeMo Guardrails」](#)の自動車向けに最適化された実装を開発し、セレンス搭載システムによる車内でのやりとりのニュアンス理解を実現
- NVIDIA DRIVE AGX OrinでCaLLM Edgeにエージェントアーキテクチャを導入して最適化を図り、次世代の車内ユーザーエクスペリエンスの向上に寄与

NVIDIAとの提携拡大により、セレンスは、自動車メーカーのお客様と協力して次世代ユーザーエクスペリエンスを開発するための、拡張性と信頼性に優れたツールやリソースを得られます。これにより、高度な性能、遅延の低下、プライバシーとセキュリティの強化、悪意のあるインタラクションや不要なインタラクションに対する強固な保護がもたらされ、充実したドライバーエクスペリエンスの提供が可能になります。

セレンスの製品技術エグゼクティブ・バイスプレジデント、ニールス・シャンツ(Nils Schanz)は次のように述べています。「当社の言語モデルファミリーであるCaLLMの性能を最適化することで、ドライバー向けに生成AI搭載ソリューションの迅速な導入を目指している自動車メーカーのお客様のコスト削減と性能向上を実現できます。CaLLMを基盤として次世代プラットフォームを進化させることで、高度な機能を通じてより迅速かつ信頼性の高いインタラクションをドライバーに提供でき、運転中の安全性、楽しさ、生産性が向上します」

NVIDIAのオートモーティブ担当バイスプレジデント、リシ・ダル氏(Rishi Dhall)は次のように述べています。「大規模言語モデル(LLM)は膨大な新しいユーザーエクスペリエンスをもたらしていますが、その規模や導入の複雑さから、開発者からエンドユーザーへ、AI搭載ソリューションの提供が困難な場合があります。今回の提携拡大により、セレンスは、NVIDIAの先進的なAIと高速化したコンピューティング技術を導入し、LLMの開発と導入を最適化できます」

セレンスの詳細については、当社の日本語ウェブサイト<https://www.cerence.com/ja>をご覧ください。最新情報については公式[LinkedIn](#)からもご覧いただけます。

### 【セレンスについて】

セレンス(Cerence Inc.)は、自動車や交通機関においてAIを活用した直感的でシームレスな体験を提供するグローバルリーダーです。数十年にわたる音声認識の専門知識とイノベーション、および生成AI、大規模言語モデル(LLM)を活用し、ドライバーと同乗者の双方にとって、より安全で、よりコネクタされ、より楽しい移動手段を実現する統合的なエクスペリエンスをお届けします。当社の技術が搭載された自動車は全世界で5億台を超え、大手自動車メーカー、交通機関OEM、テクノロジー企業と提携しながら次世代のユーザーエクスペリエンスを推進しています。マサチューセッツ州バーリントンに本社を置く当社は、世界の国と地域で事業を展開し、AIイノベーションの可能性を押し広げることに専念しています。

詳細については、当社の日本語ウェブサイトをご覧ください。<https://www.cerence.com/ja>