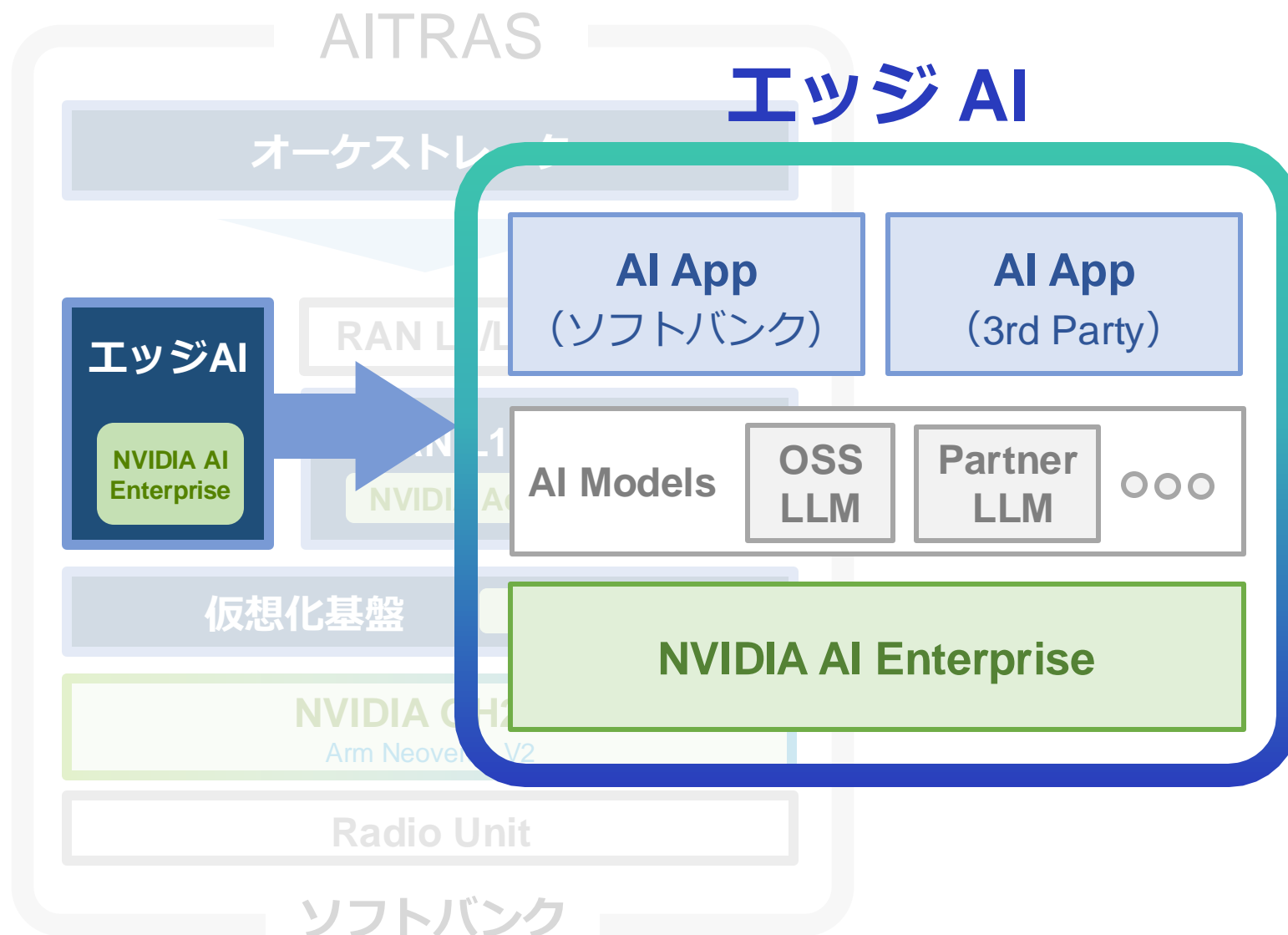


RAG Menu @Edge

AITRASの要素技術：AI Edge



オリジナル AI サービス

- 自動運転監視アプリ
- クラウドロボット
- **RAG Menu @edge**

NVIDIA Serverless API

空きリソースを
顧客需要にスポット割当

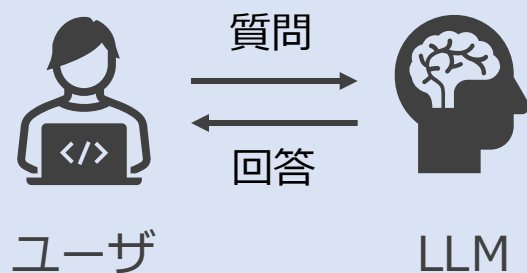
NVIDIA AI Enterprise

業界標準の
AI フレームワーク

RAG (Retrieval-Augmented Generation) とは

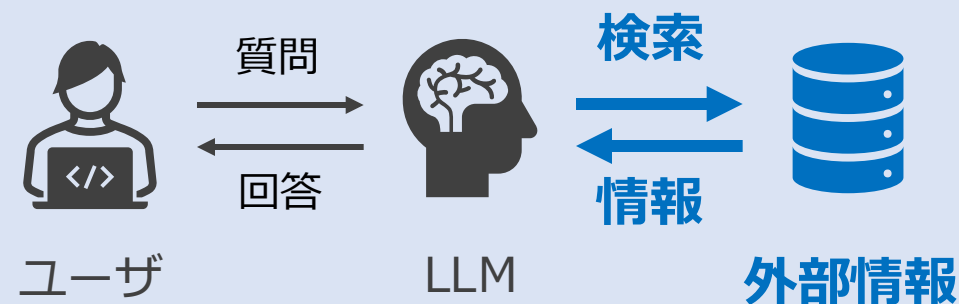
社内の機密情報など、LLM が知らない情報を回答可能

✓ 一般的な LLM



事前学習した情報のみ
に基づき回答

✓ RAG



事前学習した情報 + 外部情報
に基づき回答

マルチモーダル RAG

画像や音声など、様々なデータを RAG で扱えるように

視覚情報



例：社内資料中の図やグラフ

音声情報

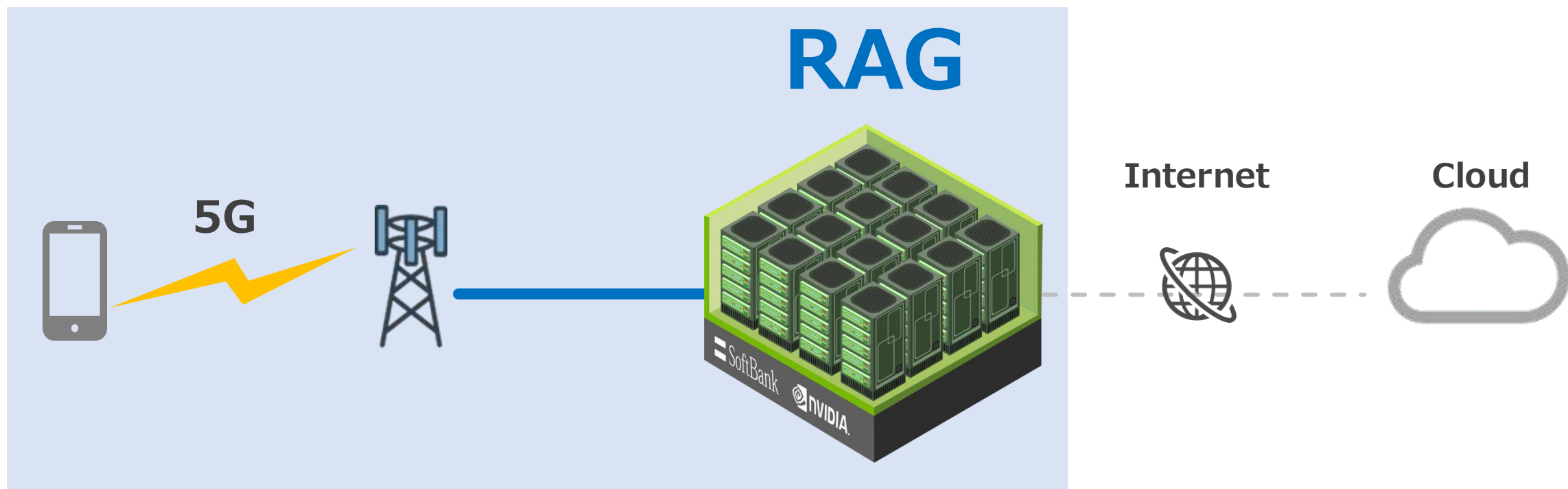


例：会議、コールセンター

エッジ上での RAG 提供

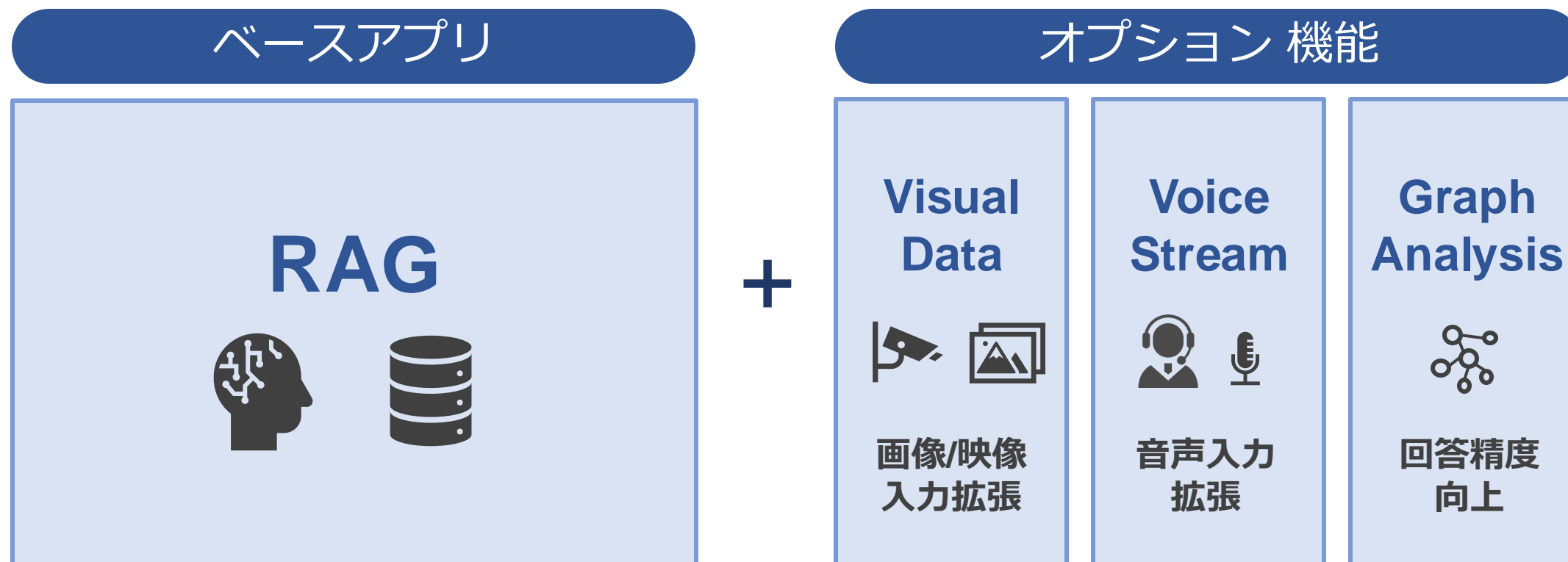
RAN はクローズドネットワーク

→ 機密情報を用いた RAG のセキュリティ確保に有効



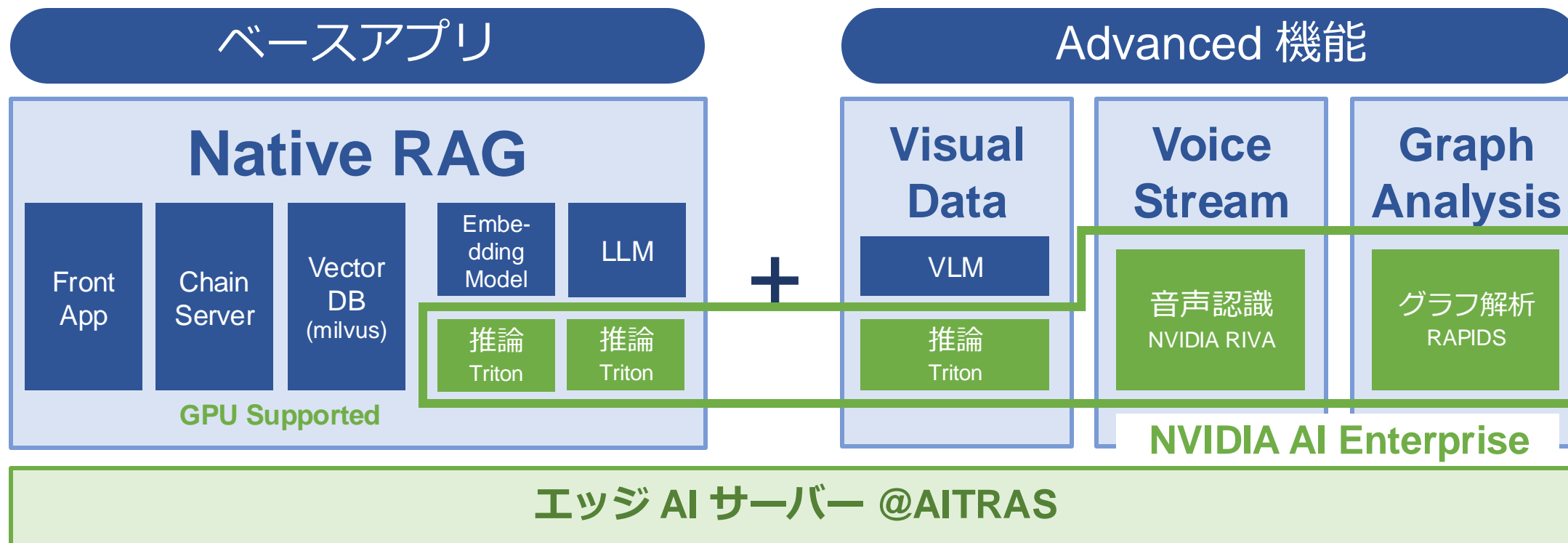
RAG Menu @Edge

顧客のニーズに応じた追加オプションを用意

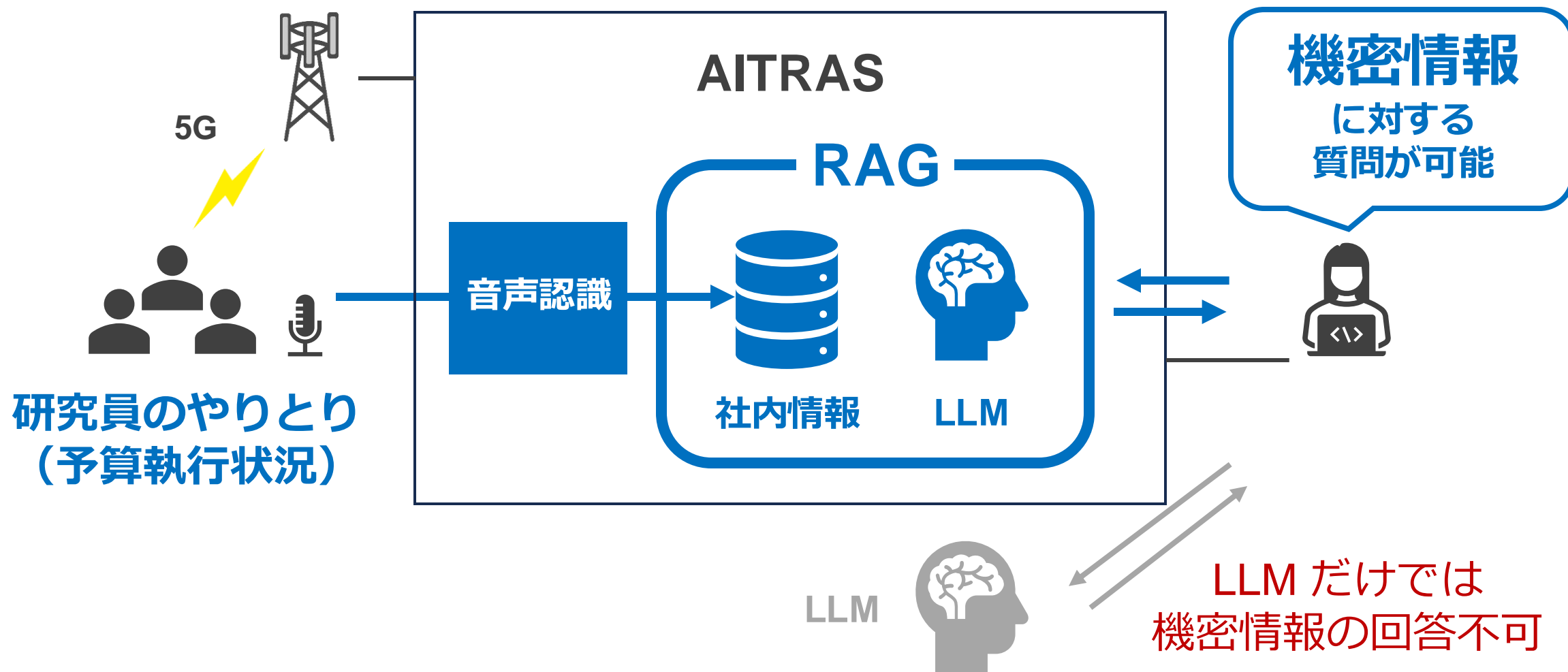


提供機能

NVIDIA AI Enterprise で実装された GPU 最適な機能群を提供



デモ：先端技術研究所 RAG




RAG DEMO

オーディオファイルを選択

58_Sentan_Budget_Status.wav

Play

  Input the audio data into the RAG

Clear



LLM Chatbox

RAG

テキストを入力してENTERを押してください

例：エンタープライズシナリオ RAG による工場オペレーション効率化

