

クラウド (LLM) ロボット

AITRASのシステム構成

AITRAS

オーケストレーター

エッジAI
NVIDIA AI Enterprise

RAN L2/L3 ソフトウェア

RAN L1 ソフトウェア
NVIDIA AI Aerial

仮想化基盤

MIG/MPS

NVIDIA GH200
Arm Neoverse V2

Radio Unit

ソフトバンクが開発した要素技術

管理・制御ソフト

無線信号の処理ソフトを高度化

仮想化基盤を構築・実装

Edge AIのサービスメニューを提供

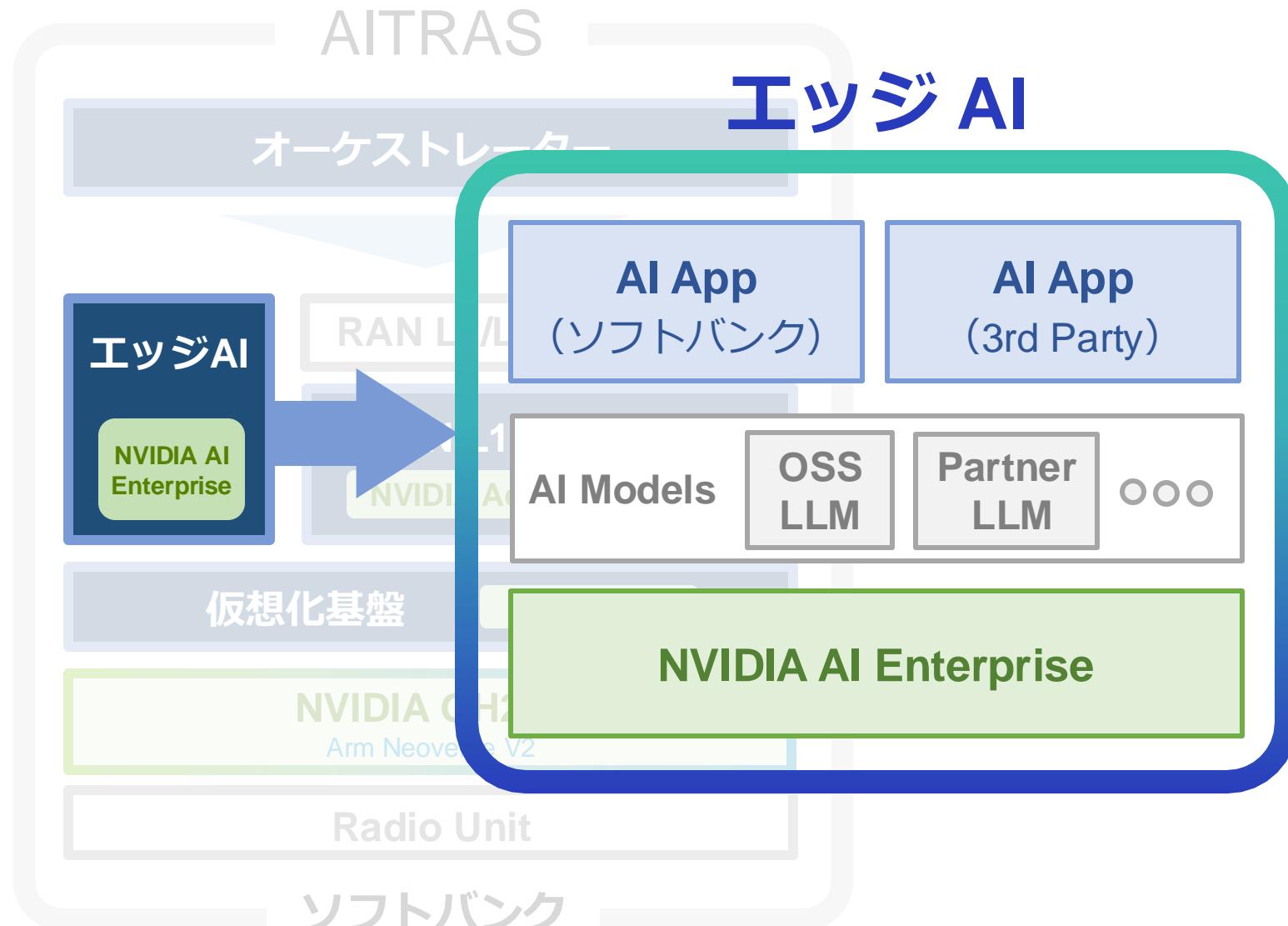
LLM
Robot

マルモーダル
LLM

高機能
RAG

NVIDIA
Serverless
API*

このデモの位置づけ



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 等

NVIDIA Serverless API @Edge

空きリソースを
顧客需要にスポット割当

NVIDIA AI Enterprise @AITRAS

業界標準の
AI フレームワーク

ロボットとの共存社会



高度な判断をする様々なロボットが活躍

処理オフロード

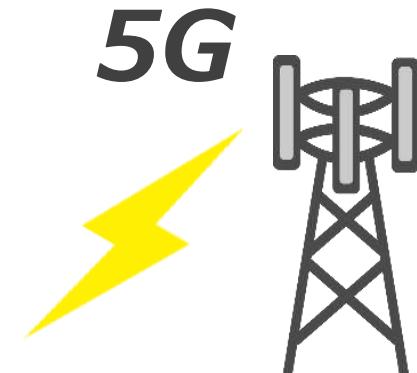
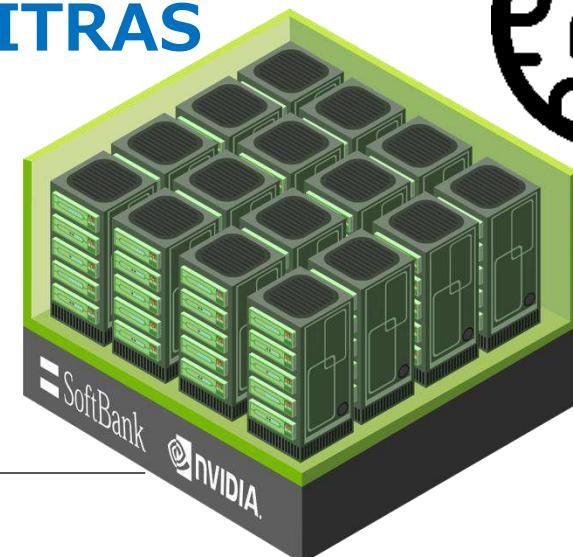
高度な判断には高度な計算機が必要

計算リソース
小



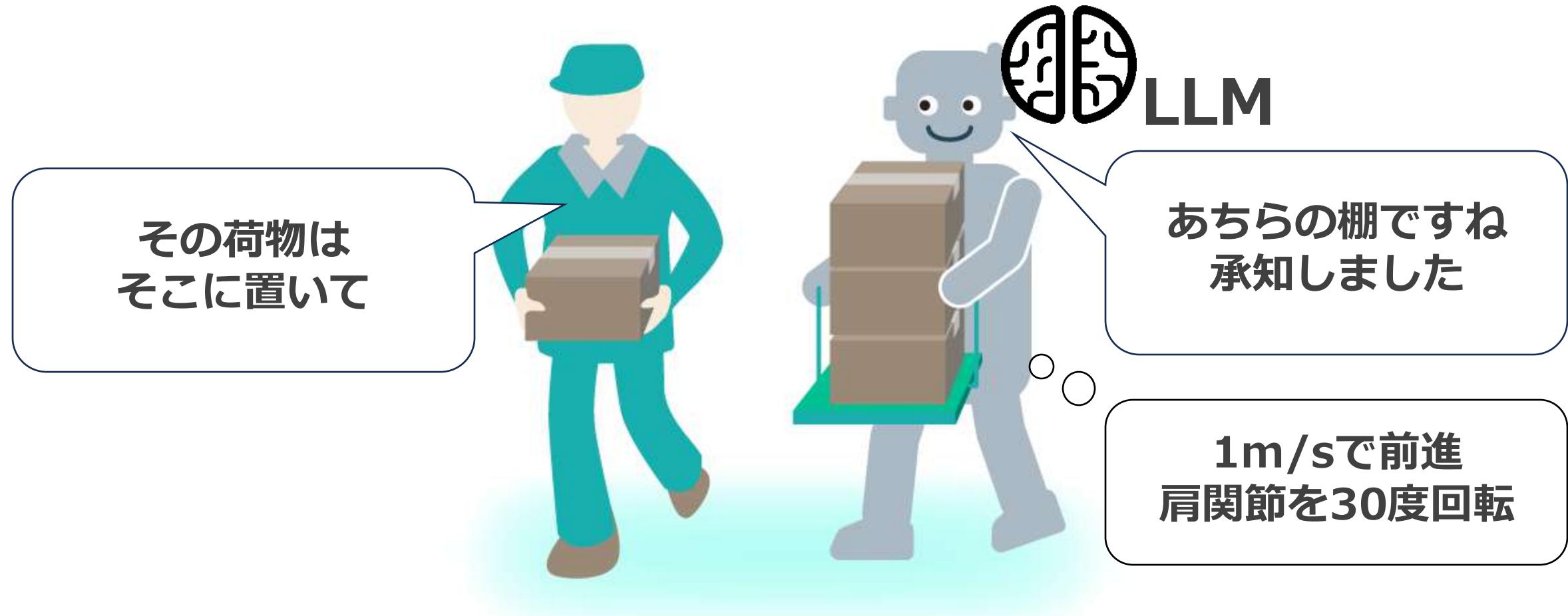
計算リソース
大

AITRAS



AITRASの方がより高度な計算機をえる

LLMロボットとは？



初めての事象にも常識をもって対応できる

ロボットを制御するには高速化が必須

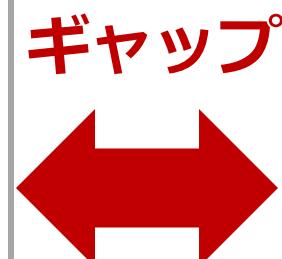
普通のLLMの遅延
質問から回答の時間



Internet


1秒以上

ロボットの制御遅延
制御情報の出力

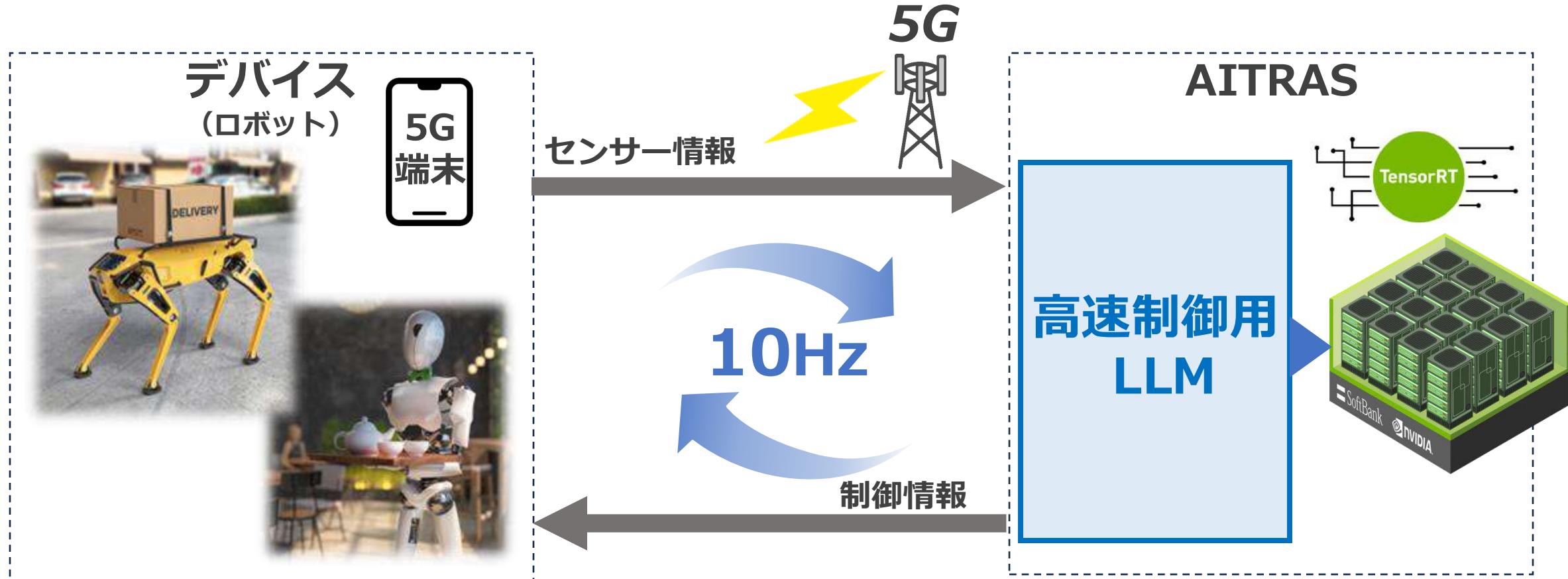


内部計算機



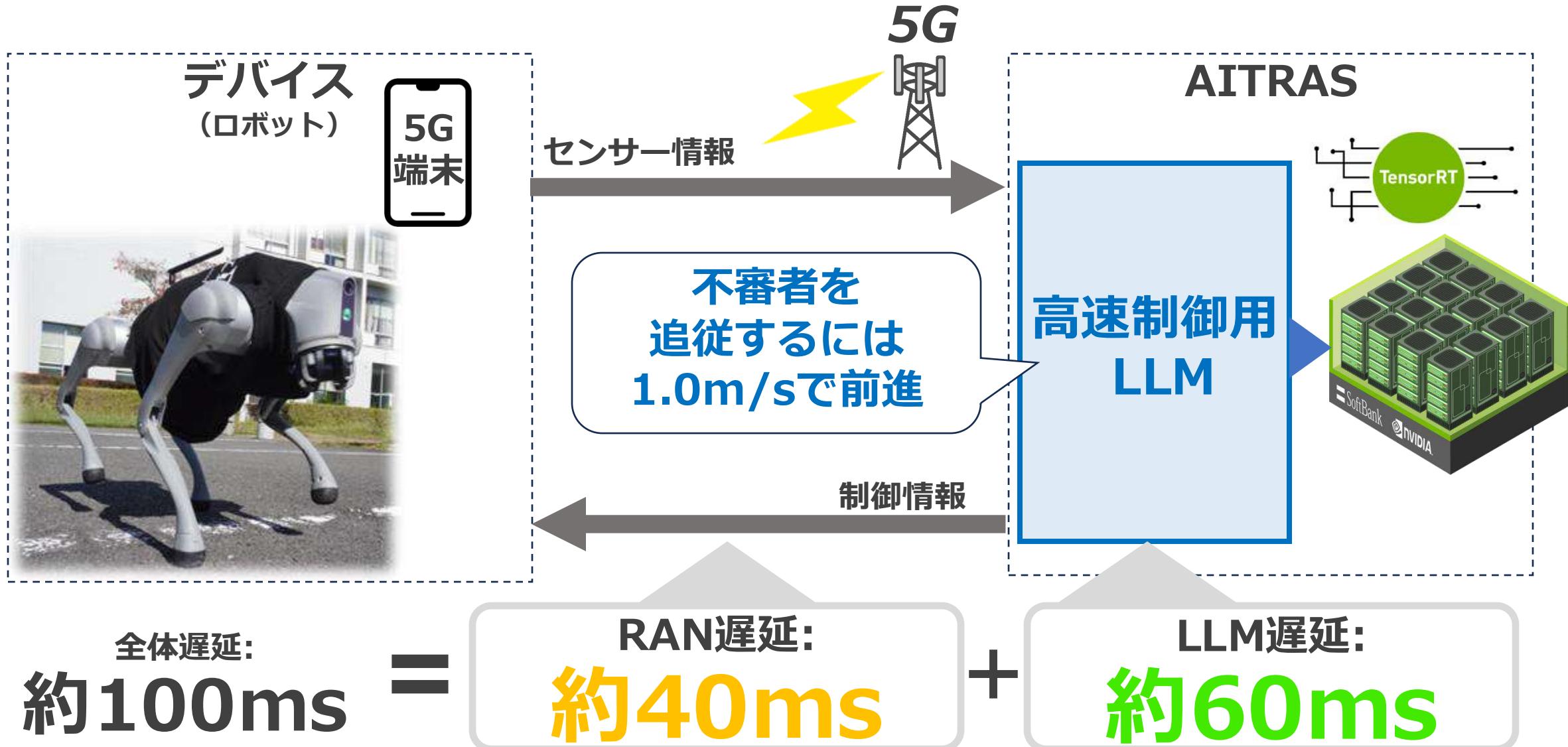

0.1秒程度

低遅延で動作するLLMロボット



LLMにより1秒間に10回のロボット制御

低遅延で動作するLLMロボット





低遲延



高遲延

