

＝SoftBank R&D

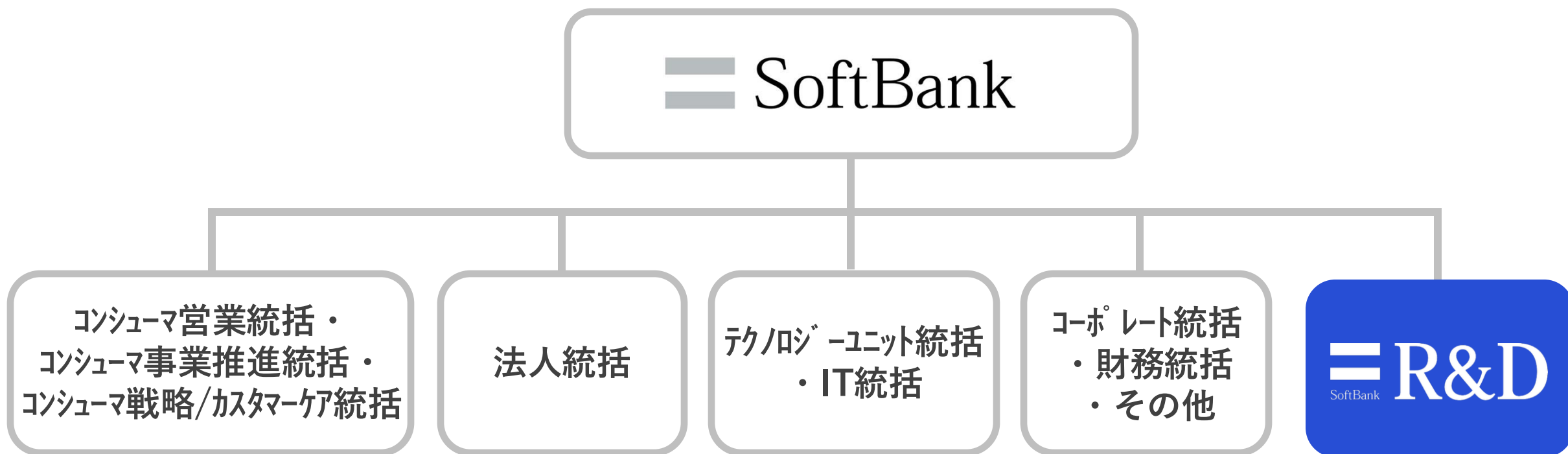
# ソフトバンクのAIインフラ戦略

2025.12.11

ソフトバンク株式会社  
執行役員 兼 先端技術研究所 所長  
湧川 隆次

# 先端技術研究所について

通信のさらなる進化や先端技術を活用した新規ビジネスのためのR&Dを中心に活動



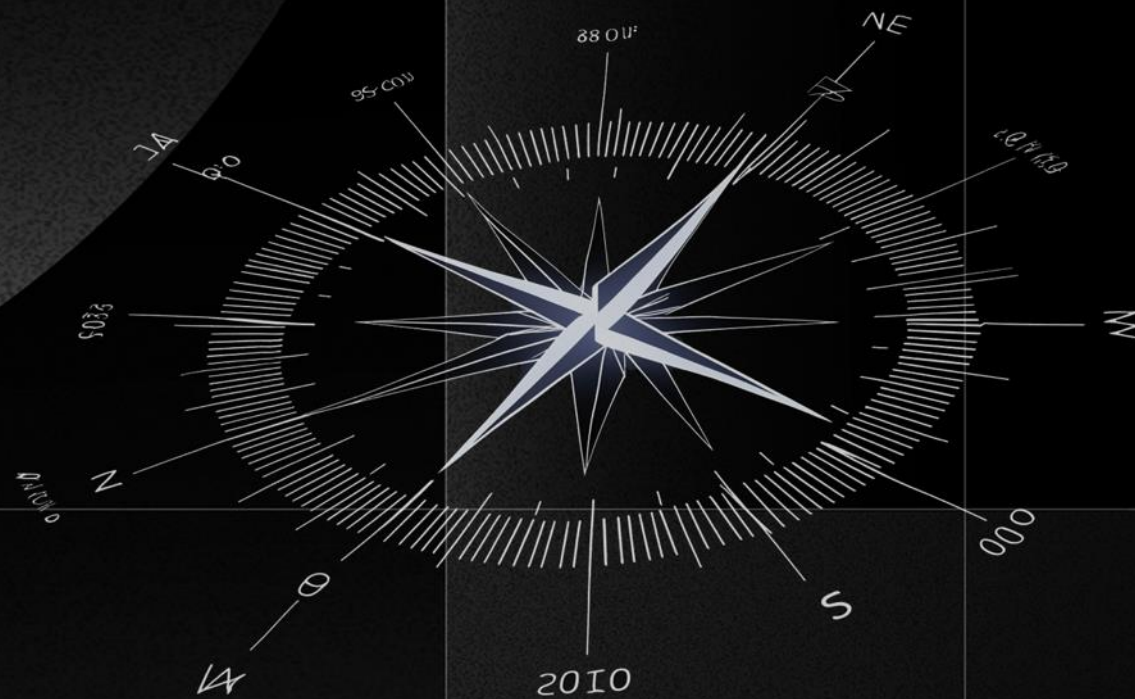
先端技術研究所  
2022年4月新設



# = SoftBank R&D

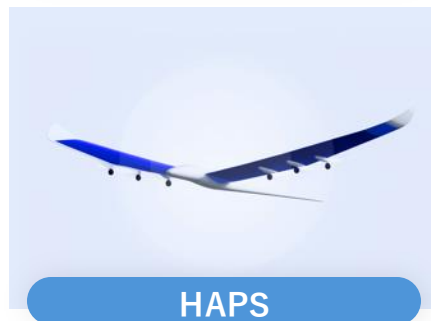
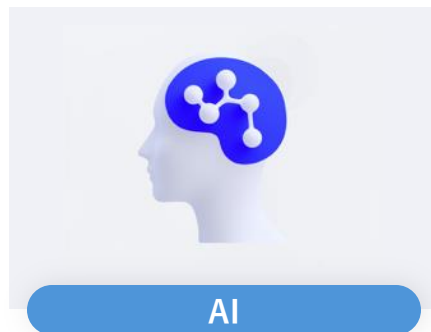
# “技術は羅針盤”

# 技術に裏付けられた ビジョンをつくる



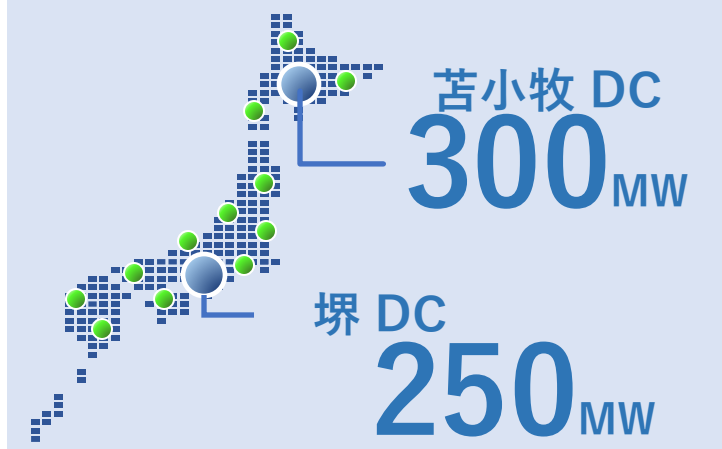
# 先端技術研究所の活動領域

70以上のプロジェクトを推進中



# ソフトバンクのAI関連活動

## AI Data Center



## AI-RAN: AITRAS



2024年8月

4x4  
on-air



2025年10月

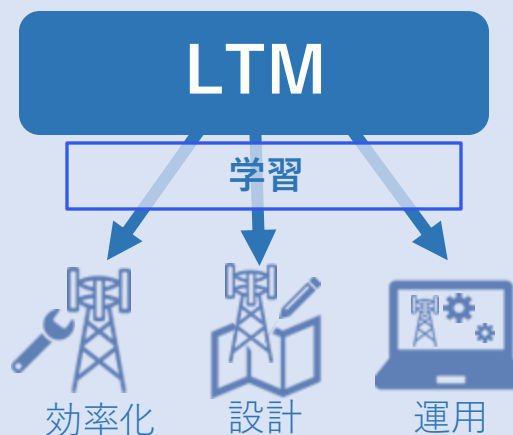
Massive MIMO

16<sub>Layers</sub>  
on-air

## LTM: Large Telecom Model



継続学習  
学習データ(~10TB)  
・SB NW設定.  
・SB KPI データ





# インフラの構造的課題を抜本的に解決

現在



データ処理/電力消費が  
都市部に集中



次世代社会インフラ



データ処理/電力を  
地方分散処理

# 分散型AIデータセンター

## Regional Brain



計算基盤

## Brain DataCenter

HPC・AI基盤

Data Lake

量子  
コンピュータ



- ・ 計算基盤を分散配置
- ・ グリーンエネルギー調達



# AIコンピューティング基盤

4,000基超のNVIDIA Blackwell GPUを搭載



## 「NVIDIA Blackwell GPU」を搭載した 「NVIDIA DGX SuperPOD」として、 世界最大のAI計算基盤を構築

～4,000基超の「NVIDIA Blackwell GPU」の整備が完了～

2025年7月23日  
ソフトバンク株式会社

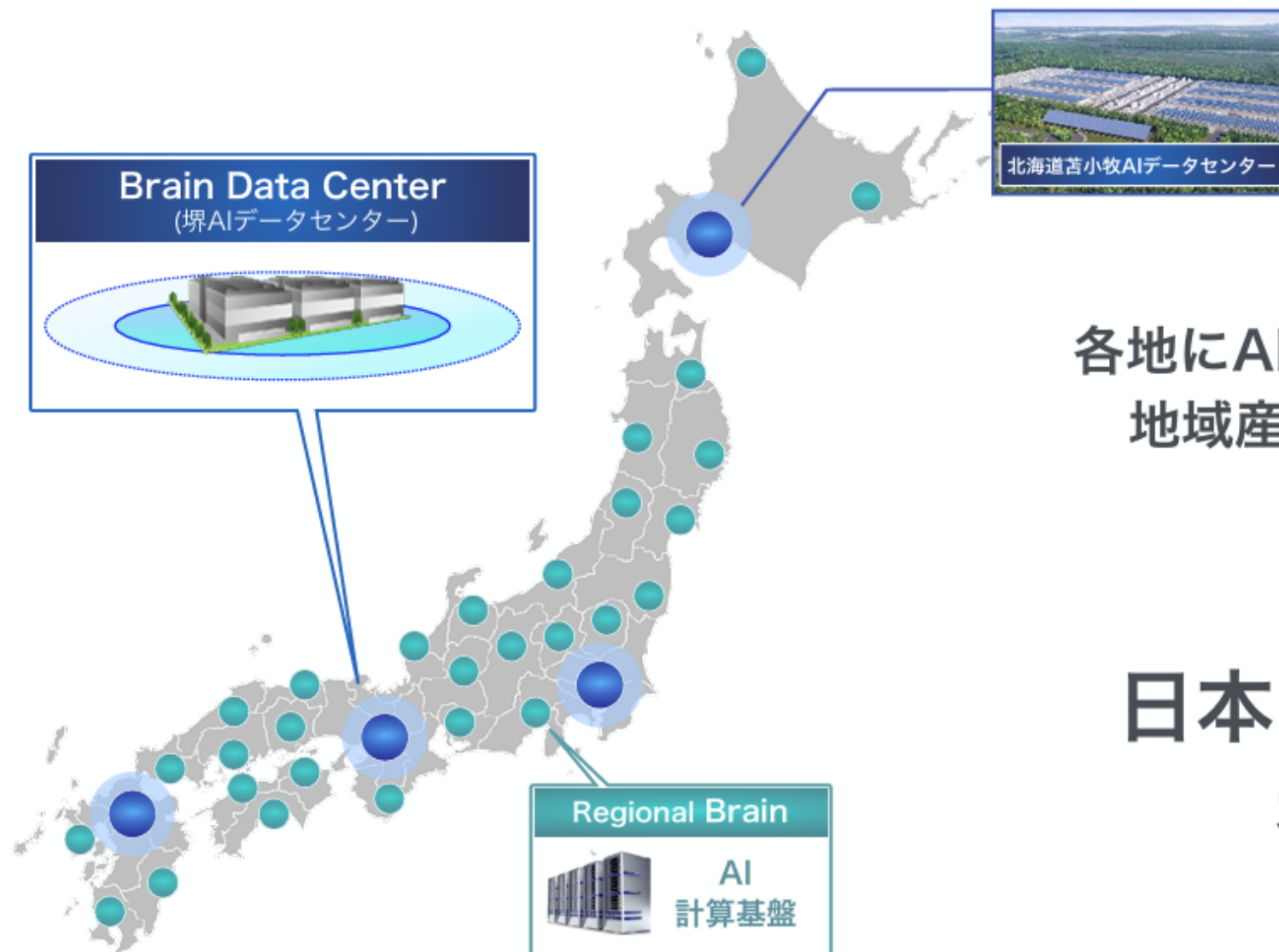
ソフトバンク株式会社（以下「ソフトバンク」）は、現在拡張を進めているAI（人工知能）計算基盤において、4,000基超の「NVIDIA Blackwell GPU」を搭載した、「NVIDIA DGX B200」システムで構成された「NVIDIA DGX SuperPOD」の構築を2025年7月22日に完了しました。このAI計算基盤は、「NVIDIA DGX B200」システムで構成された「NVIDIA DGX SuperPOD」として世界最大<sup>※1</sup>の計算能力を実現しています。また、この計算基盤は、AI処理に必要な大量のデータを高速かつ安定してやり取りできるよう、NVIDIAの最新ネットワーク技術「Quantum-2 InfiniBand」を採用しています。さらに、企業向けのAI開発や運用に必要なツール群「NVIDIA AI Enterprise」にも対応しており、AIの開発や運用をより効率的かつ安全に行える環境を提供します。これにより、ソフトバンクが構築したAI計算基盤全体のGPU（Graphics Processing Unit）は1万基超となり、計算処理能力の合計は13.7EFLOPS（エクサフロップス<sup>※2</sup>）となりました。

出典：2025/7/23 当社プレスリリース

“「NVIDIA Blackwell GPU」を搭載した「NVIDIA DGX SuperPOD」として、世界最大のAI計算基盤を構築”



# AIデータセンターを中心とした「産業集積地」構想

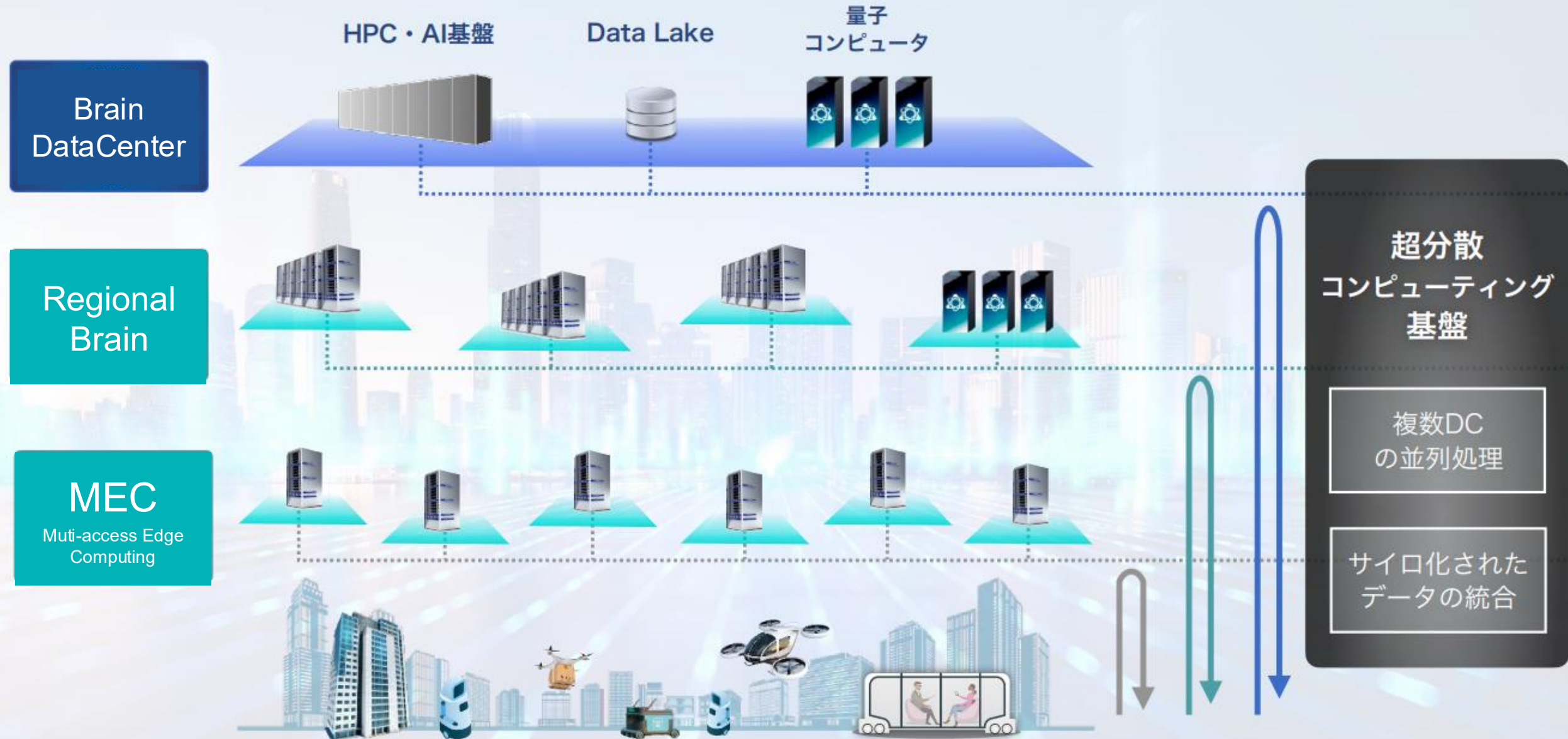


各地にAIデータセンターを構築し  
地域産業・経済の成長を牽引



日本の産業を支える  
社会基盤へ

# 分散型AIデータセンターの配置

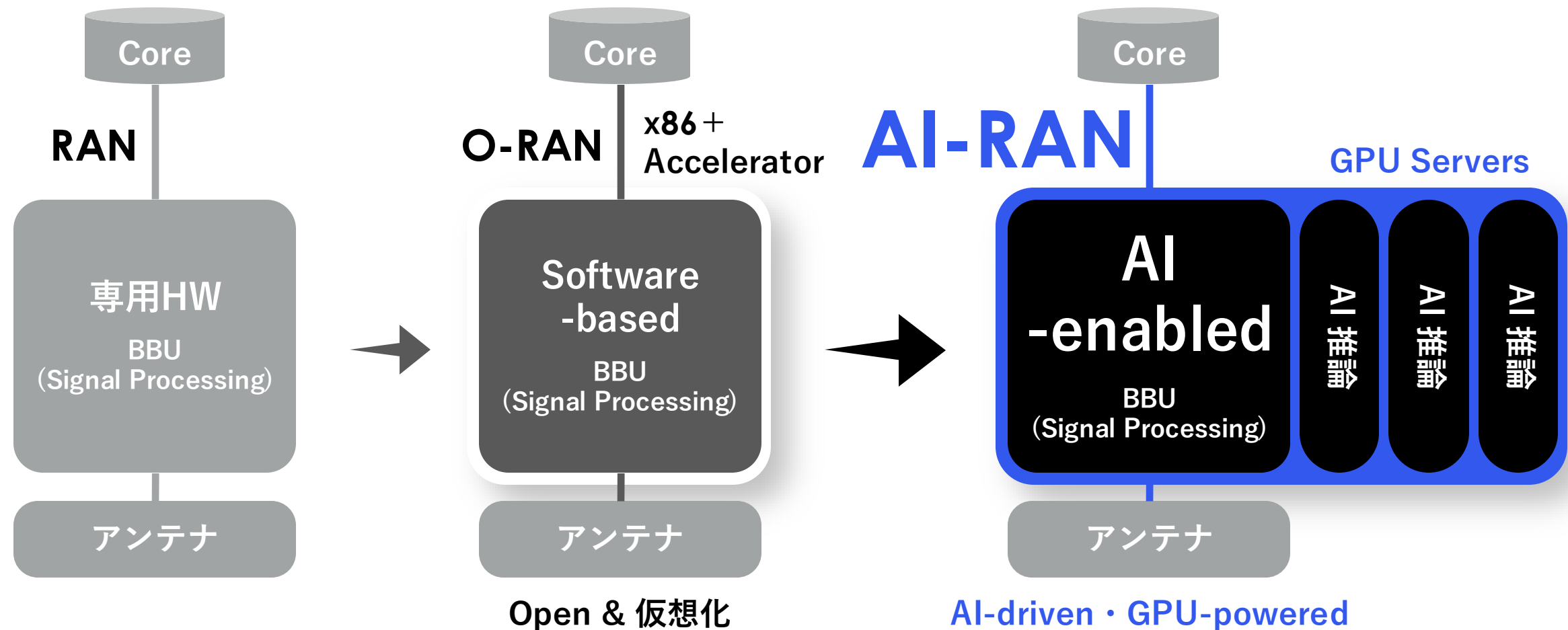






# AI-RAN

# AI-RANとは？





# AI-RAN解説動画

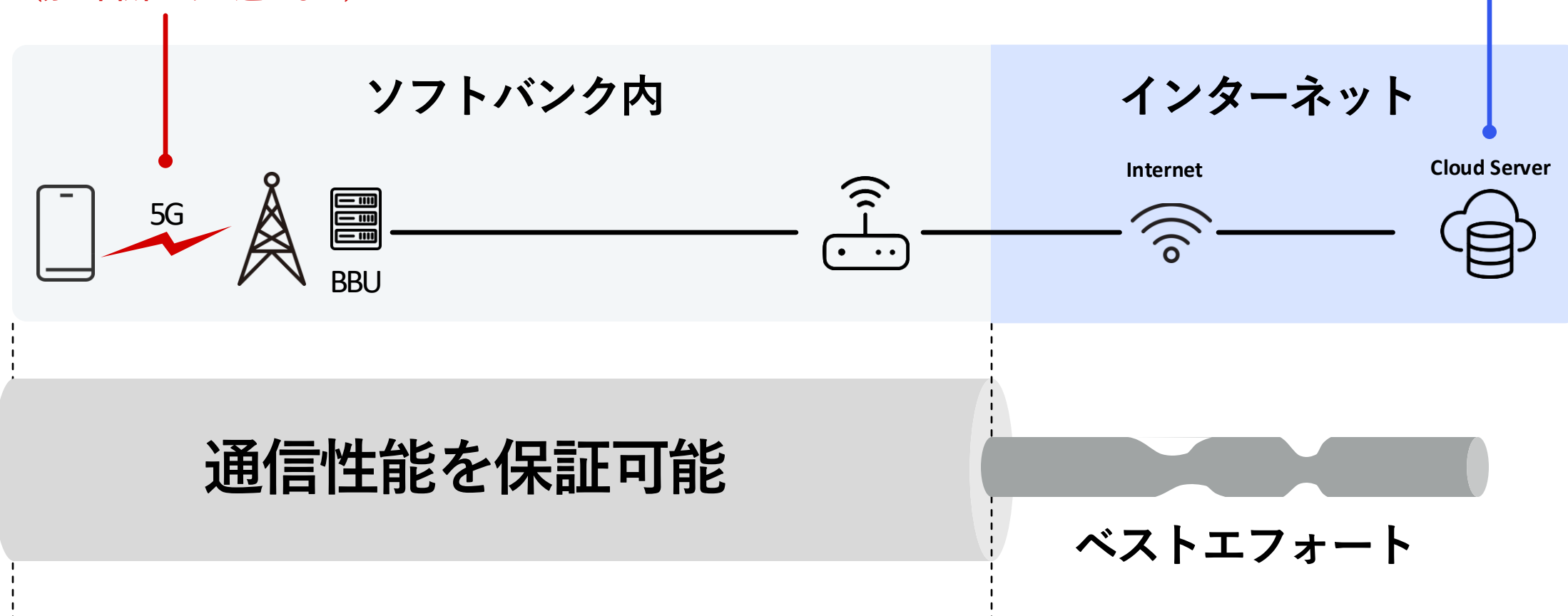


<https://www.youtube.com/watch?v=4cY2bO9xz0Q>

# AIサービスを使用するには

**5Gで高速化**  
(赤い箇所だけが速くなる)

サービスがあるのはここ

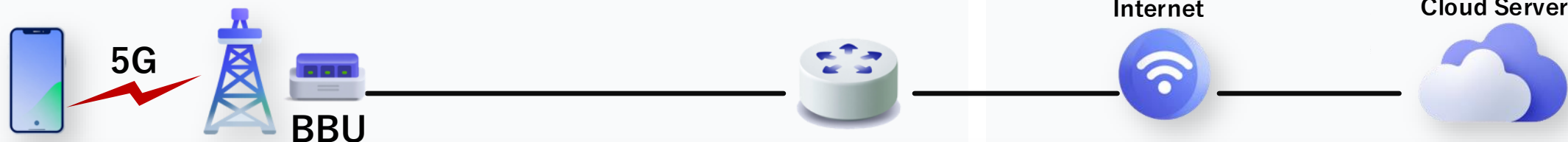




# キャリアの強みを最大限に

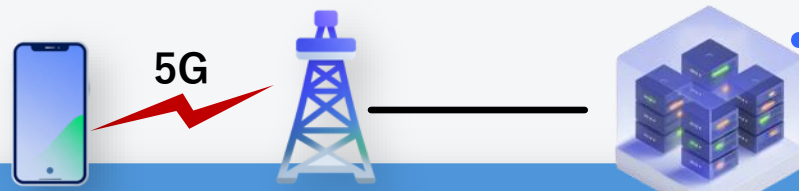
通信 (キャリアグレード)

計算資源 (Best Effort)



200+ msec

キャリアグレードの「通信 + 計算資源」



~20 msec



## AI-RAN

- vRAN
- AI 推論

# AI-RANで進化するAIサービス



低遅延

高帯域通信

高セキュリティ

分散コンピューティング

AI推論



ロボットとの  
リアルタイムなやりとり



手術補助



自律型ロボット  
意思決定



機密情報  
LLM RAG



# NVIDIAとのGPUに関する取り組み

## 2019年からGPU-vRANに関する検証を共同で実施

GPU vRAN参戦

GPU vRAN検証

MEC AI検証

AI-on-5G Lab開設

AI-RAN Alliance 設立

2019

衝撃の「NVIDIA、通信業界に“再参入”」。KDDIとSBが採用表明した“ソフトウェア化する5G基地局”とは

…「EGX Edge Supercomputing Platform」という新しい“エッジコンピューティング”と呼ばれる最新技術向けのサーバー製品群を発表した。また、これを活用した5G向けの通信機器を、スウェーデンのエリクソン、米国のレッドハットと協業して提供していくと明らかにした。

しかも、これらの製品は、日本の通信キャリアであるKDDI、ソフトバンクが評価を開始していると明らかにした。…

出典：TECH INSIDERの記事を基に作成

2020

NVIDIAのGPUを活用した5G仮想基地局の技術検証を実施

～処理速度や消費電力などで優れた性能を確認～

ソフトバンク株式会社（以下「ソフトバンク」）は、NVIDIA Corporation（以下「NVIDIA」）と協同で、NVIDIAのGPU（Graphics Processing Unit）を活用して構築した5G仮想基地局（vRAN）の技術検証を実施した。検証の結果、処理速度や消費電力などで優れた性能を確認した。

5Gは、高速・大容量通信に際して、処理速度を向上させることが重要である。ソフトバンクは、NVIDIAのGPUを活用した5G仮想基地局（vRAN）の技術検証を実施した。検証の結果、処理速度や消費電力などで優れた性能を確認した。

検証の結果、処理速度や消費電力などで優れた性能を確認した。また、NVIDIAのGPUを活用した5G仮想基地局（vRAN）の技術検証を実施した。検証の結果、処理速度や消費電力などで優れた性能を確認した。

検証の結果、処理速度や消費電力などで優れた性能を確認した。また、NVIDIAのGPUを活用した5G仮想基地局（vRAN）の技術検証を実施した。検証の結果、処理速度や消費電力などで優れた性能を確認した。

検証の結果、処理速度や消費電力などで優れた性能を確認した。また、NVIDIAのGPUを活用した5G仮想基地局（vRAN）の技術検証を実施した。検証の結果、処理速度や消費電力などで優れた性能を確認した。

2021

5GとMEC、「NVIDIA Maxine」を活用して映像を“超解像”する実証実験に成功

～商品用ウェブ動画を活用～

ソフトバンク株式会社（以下「ソフトバンク」）は、5GとMEC（Multi-access Edge Computing）を活用して、NVIDIA Maxine（以下「Maxine」）を活用して映像を“超解像”する実証実験に成功した。検証の結果、映像の品質が向上し、商品用ウェブ動画を活用した。また、NVIDIA Maxineを活用して映像を“超解像”する実証実験に成功した。



2022

完全仮想化されたクラウド5Gの商用化に向けて研究施設「AI-on-5G Lab」をNVIDIAと共同で開設

～vRANとMEC統合したクラウド5Gの商用化を目指す～

ソフトバンク株式会社（以下「ソフトバンク」）は、NVIDIA Corporation（以下「NVIDIA」）と共同で、完全仮想化されたクラウド5Gの商用化に向けて研究施設「AI-on-5G Lab」を開設した。検証の結果、vRANとMEC統合したクラウド5Gの商用化を目指す。また、NVIDIAのGPUを活用した5G仮想基地局（vRAN）の技術検証を実施した。



2024

AI-RAN ALLIANCE



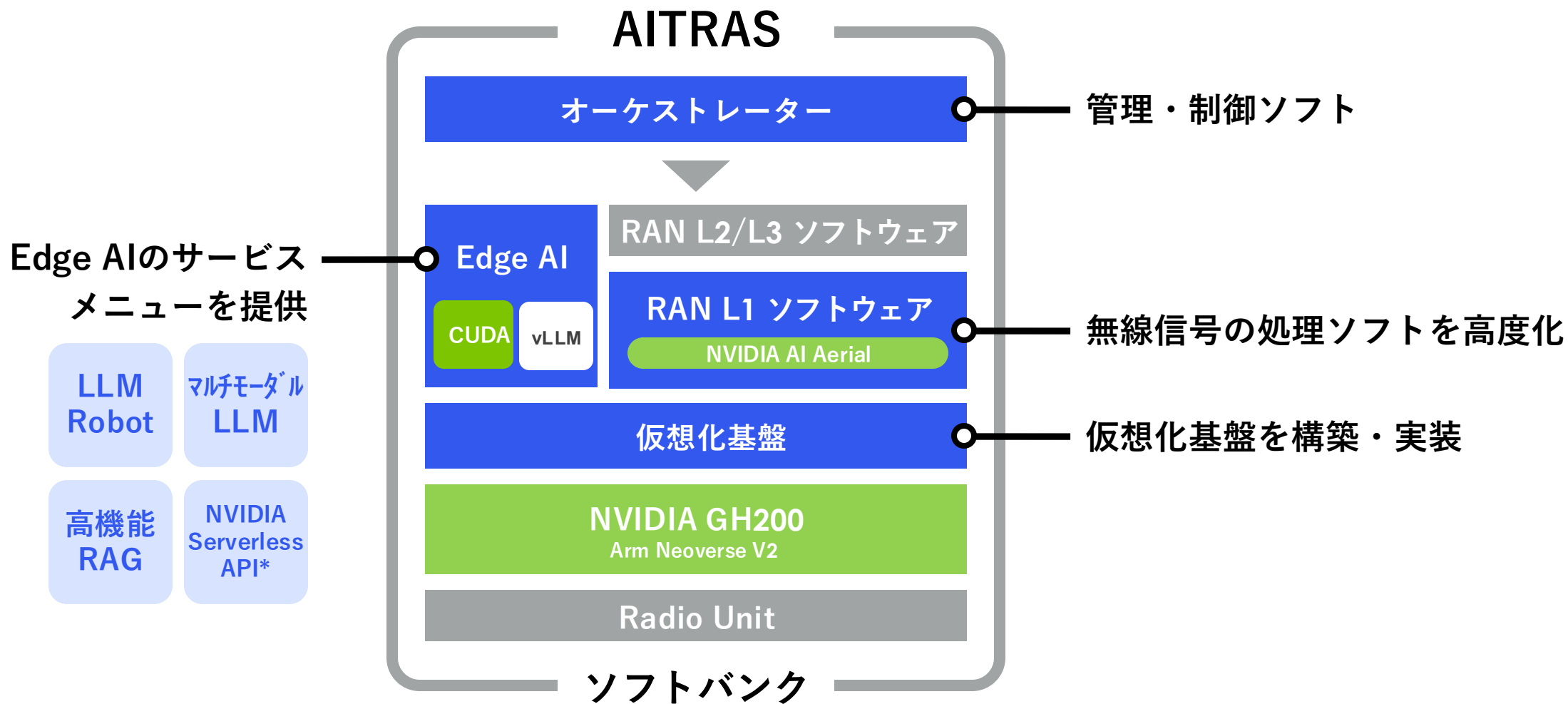
ア イ ト ラ ス

# AITRAS

AI-RANコンセプトに基づいた  
ソフトバンクオリジナルプロダクト

# AITRASのシステム構成

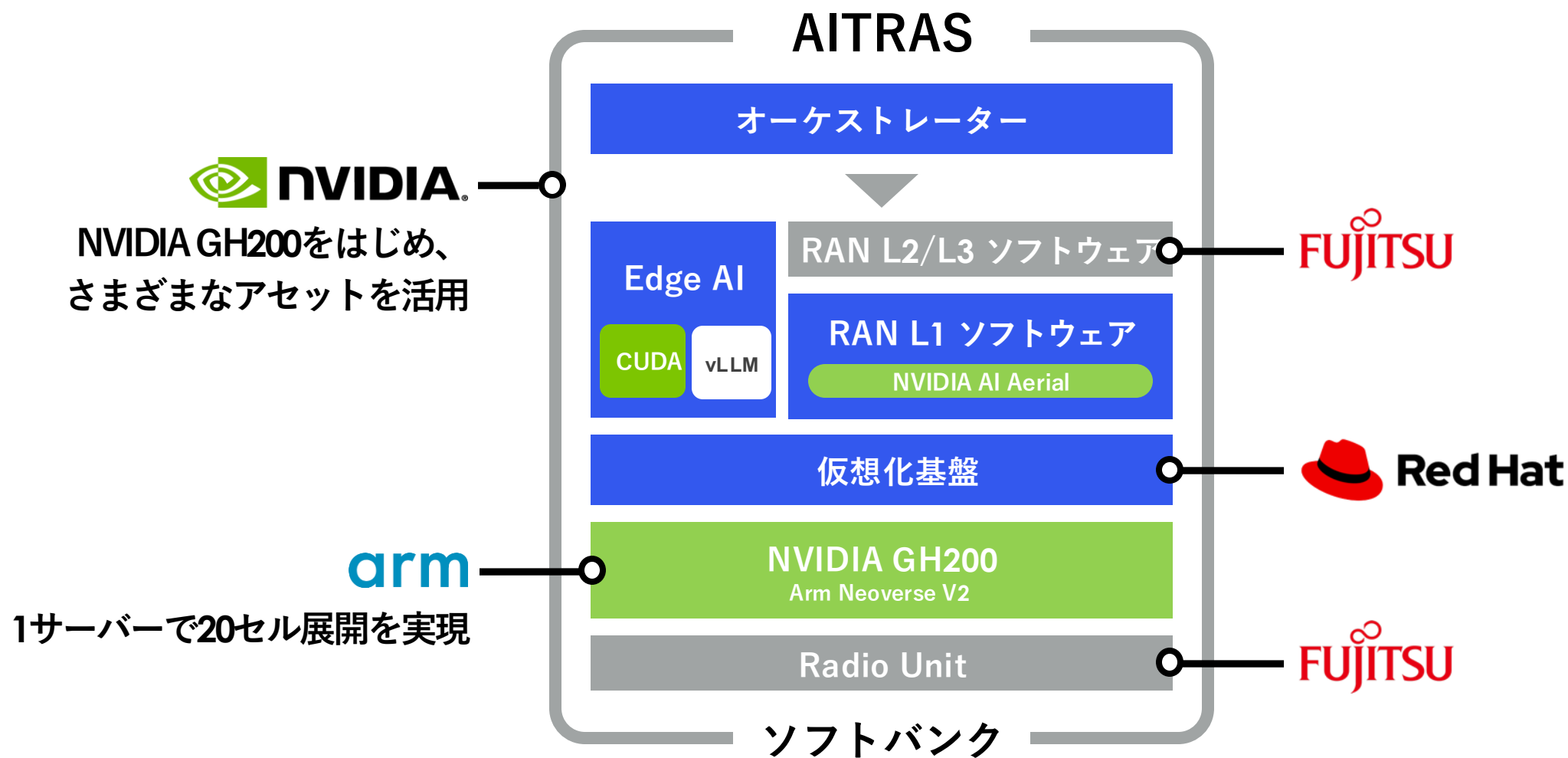
ソフトバンクが開発した要素技術





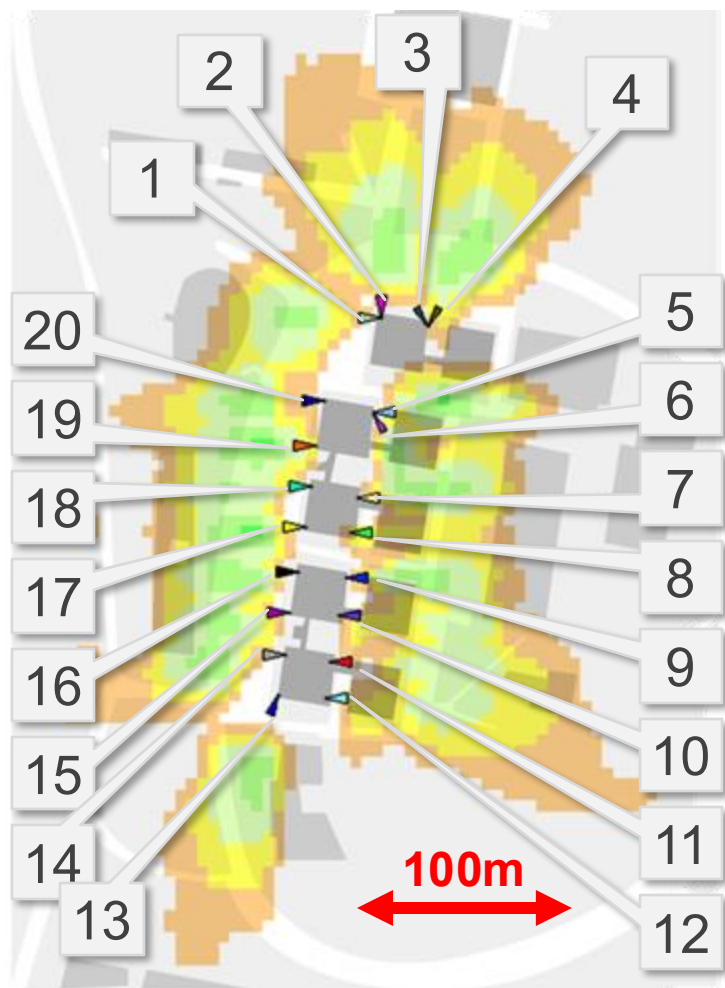
# AITRASのシステム構成

パートナーとのコラボによる高性能化



# 2024年11月@SFC AITRAS 20セルをオンエア(実験局)

SoftBank

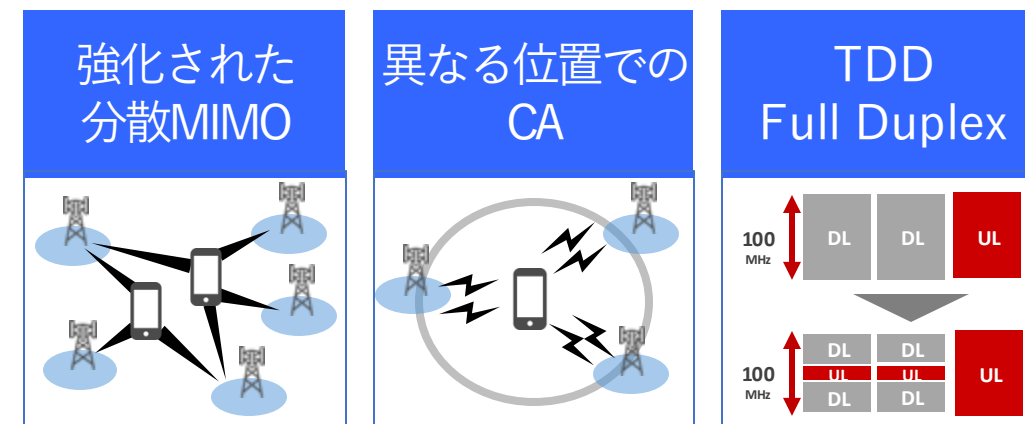
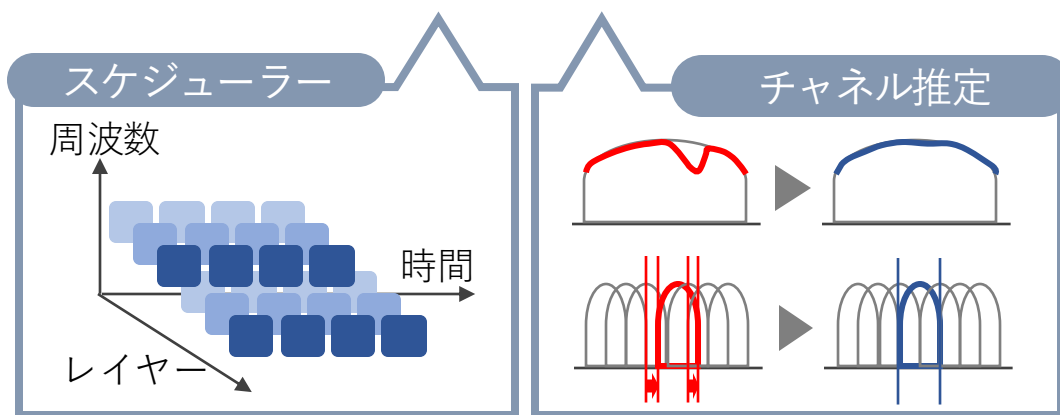
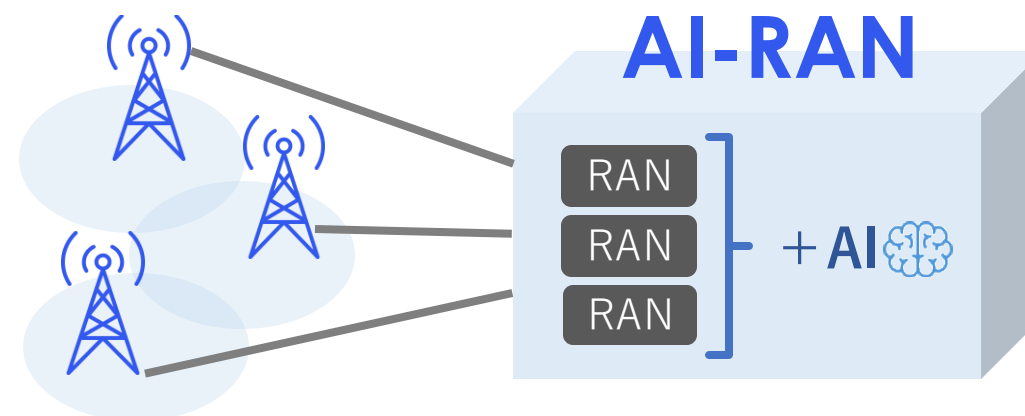


# AIによって無線リソース効率最大化へ

## AIによる単独セルの最適化



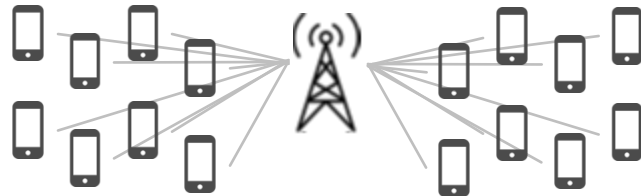
## AIによるセル間の連携・最適化





# AI無線開発実績

## MU-MIMO

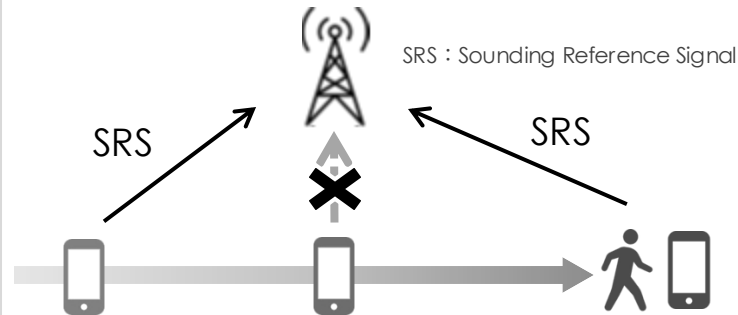


多数・多種デバイス接続による  
非効率なリソース割当

 **MACスケジューラー**

スループット向上  
**9%**

## ビームフォーミング

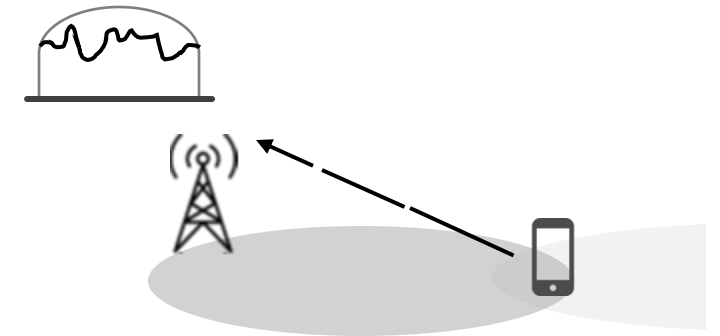


多数デバイス接続による  
ビームフォーミング性能劣化

 **SRS補完**

スループット向上  
**13%**

## UL信号処理



干渉による  
ULチャネル劣化

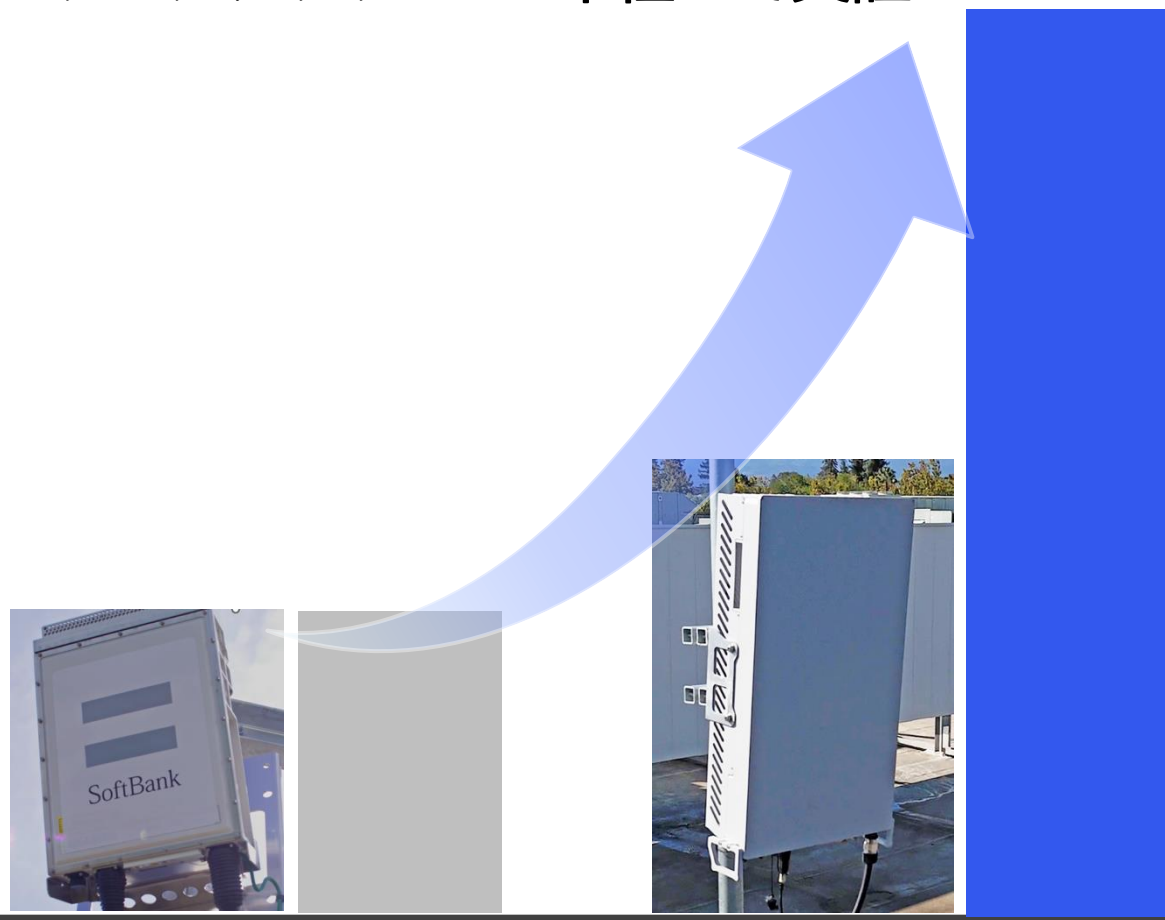
 **ULチャネル補完**

**with Transformer**

スループット向上  
**30%**

# Massive MIMO 16 Layer On-Air

サンタクララNvidia本社にて実証

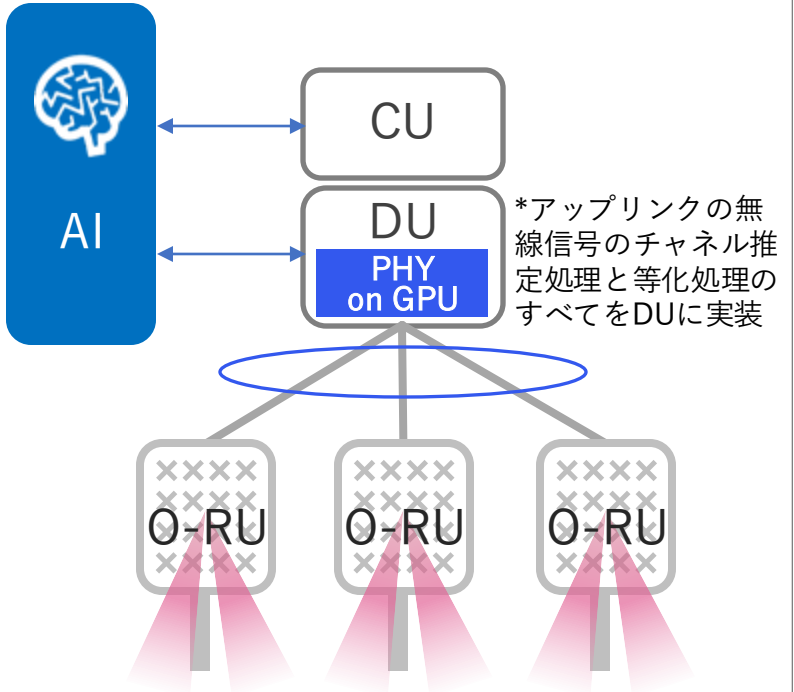
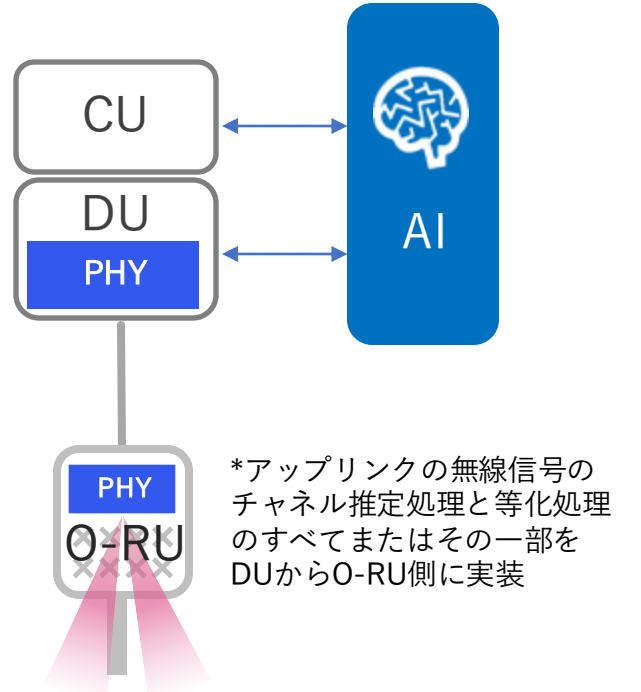


SU-MIMO

MU-MIMO

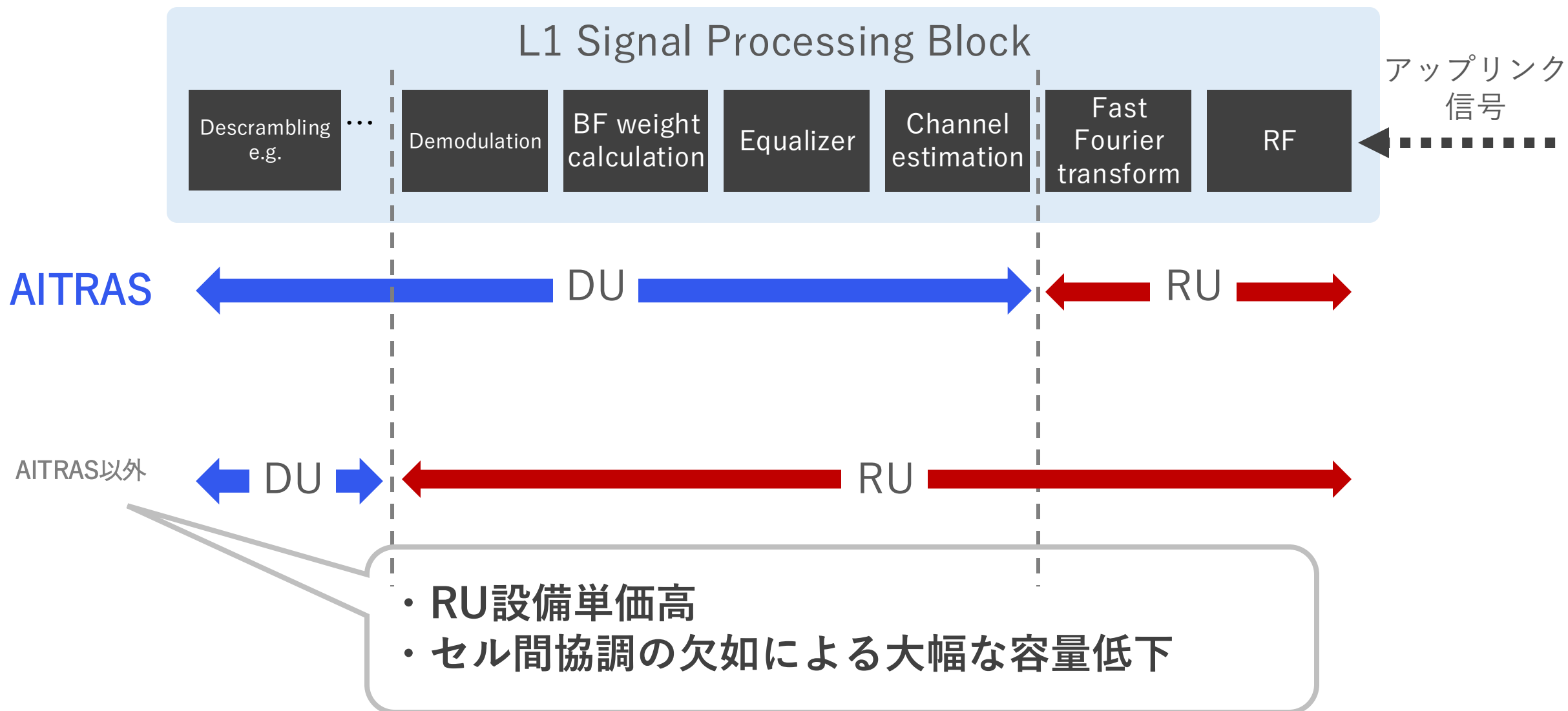
スループット  
3倍向上

# AITRASシステム構成

	AITRAS	AITRAS以外
システム構成 O-RAN Split Option 7.2x	 <p>*アップリンクの無線信号のチャンネル推定処理と等化処理のすべてをDUに実装</p>	 <p>*アップリンクの無線信号のチャンネル推定処理と等化処理のすべてまたはその一部をDUからO-RU側に実装</p>
AIで利用可能な無線品質最適化に必要な情報	完全	限定的/局所的
アップリンクチャンネル補間による無線品質の改善	高	中～低
ビームフォーミングの最適化	高	中～低



# Massive MIMOをソフトウェアで完全実装



# AI-RAN<sup>TM</sup> ALLIANCE

# 2024年2月AI-RAN アライアンス設立

共同での課題克服



- ✓ AI-RANの活用と高性能化
- ✓ 新たな収益機会の創出
- ✓ AI-RANエコシステムの構築
- ✓ 研究開発活動を通じたパフォーマンス向上



**Chair:**  
Dr. Alex Jinsung Choi,  
SoftBank

## 3つのWorking Group

### AI-for-RAN



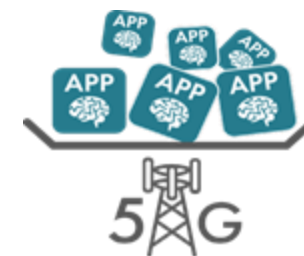
周波数の利用効率

### AI-and-RAN



設備の利用効率

### AI-on-RAN



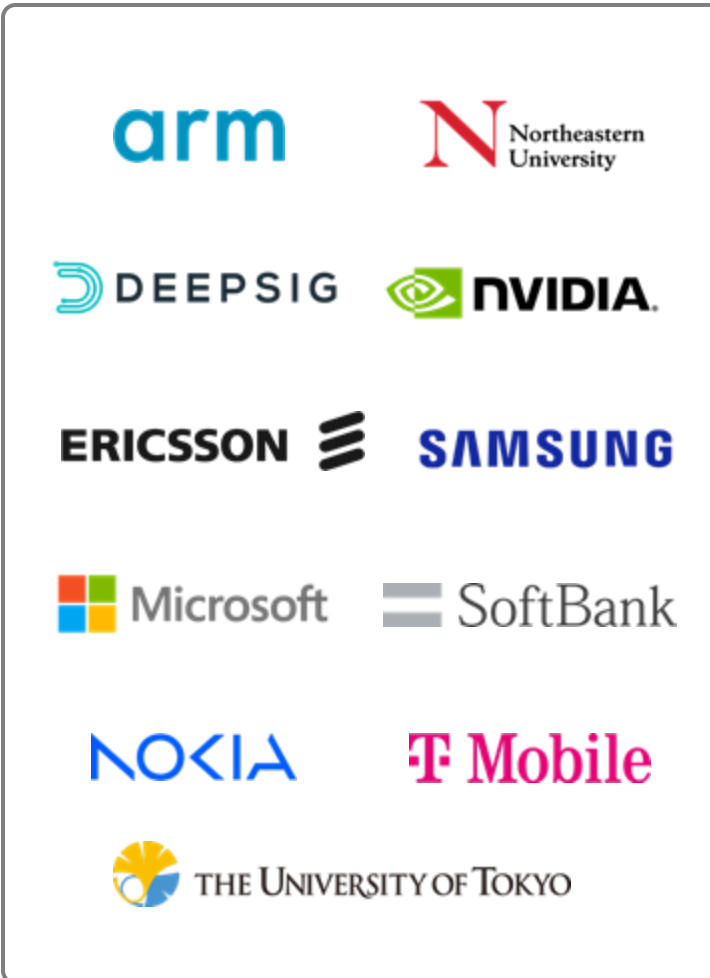
新規サービス



# AI-RAN Alliance Members: 100社を超えた



## Founding Members (11)



## General Members (104)

- 6WIND
- A10 Systems Inc. d/b/a AIRANACULUS
- Aira Technology
- AirHop Communications
- Ajou University
- Allbesmart Lda
- Altera Corporation
- Amdocs
- Amini AI
- Arizona State University
- AsiaInfo Technologies (H.K.) Limited
- Battelle Memorial Institute
- Beyond-G Global Innovation Center, Hanyang University, Seoul, Korea
- Bloxtel Inc.
- Boost Mobile Network
- Booz Allen Hamilton Inc.
- BROADRADIO INTERNATIONAL PTE. LTD.
- BubbleRAN
- CableLabs
- Capgemini America Inc
- Cisco Systems, Inc.
- Cloudera
- Cloudly, Inc.
- CLOUDRAN.AI PTE.LTD.
- Cohere Technologies
- Concordia University
- CTOne Inc.
- Dell Technologies
- EdgeCortex Inc.
- Ennoia Technologies, Inc.
- Eridan
- ETRI
- EURECOM
- F5, Inc.
- Fujitsu Limited
- G Reigns
- GlobalLogic Inc.
- Globe Telecom
- Harman Connected Services
- Hermes Network
- Hewlett Packard Enterprise
- Indian Institute of Technology Jodhpur
- Indian Institute of Technology Madras
- Indosat Ooredoo Hutchison
- Institute for Information Industry
- Itochu Techno Solutions Corporation
- ITRI
- Keio Research Institute at SFC, Keio University
- Keysight Technologies
- King's College London
- Korea University
- KT
- Kyunghee University
- LeadingSolution Inc
- LG Uplus
- LITE-ON TECHNOLOGY CORPORATION
- Matsing
- MediaTek Inc.
- Metanoia Communications Inc.
- MITRE Corporation
- Myrtle.ai
- National Yang Ming Chiao Tung University
- NEC Corporation
- NIA (National Information-society Agency)
- Paradise Mobile
- Poznan University of Technology
- Pure Storage
- Q Broadcasting LLC d/b/a Q
- Qualcomm Technologies Inc.
- Quanta Cloud Technology (QCT)
- Queen Mary University of London
- Rakuten Symphony
- Red Hat. K.K.
- RIMEDO SP. Z O.O. (Rimedo Labs)
- Rohde Schwarz GmbH Co KG
- Seoul National University
- Siemens Digital Industries Software
- Singapore University of Technology and Design
- SK Telecom
- SpectrEdge Wireless Inc
- Sungkyunkwan University System Software Lab
- Supermicro
- SynaXG Technologies Pte. Ltd.
- Synergy Design Technology Limited
- Tech Mahindra Ltd.
- Techbros GmbH
- Telecommunications Technology Association (TTA)
- Texas A&M Engineering Experiment Station
- The University of York
- Tiami Networks
- Tietoevry
- Toyota Motor Corporation
- Turkcell Iletisim Hizmetleri A.S.
- Ulsan National Institute of Science and Technology (UNIST)
- University of Bristol, Smart Internet Lab
- University of Leeds
- University of Oulu
- VIAVI Solutions
- Vodafone
- VTT Technical Research Centre of Finland Ltd.
- Wind River
- Xecurity Pulse
- Yonsei University
- zTouch Networks Inc

+ pending applications



# AI-RAN ユースケース

# 2025/12/1 安川電機との協業発表

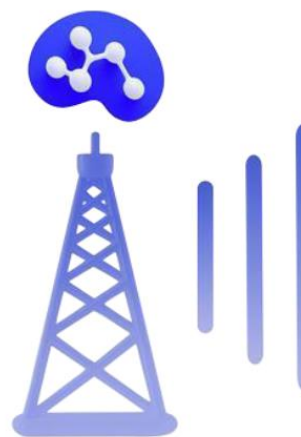
SoftBank

**YASKAWA**

SoftBank



**AIロボティクス**



**AI-RAN**



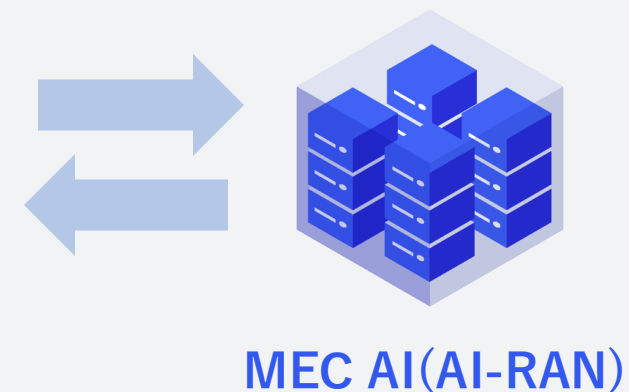
# AI-RANによるスマートビルの実現

AI-RAN



# フィジカルAIデモ

iREX 2025にてAI-RANとロボット、ビルOSが連携したデモを展示



# キャリア品質の通信 × エッジAI



新サービス



新たなアーキテクチャー



すべてのエンタープライズへ新たな機会を提供



