

# 進化するAIを支えるFUJITSU-MONAKA — AIサロゲートとAgentic AIの展望 —

2026年9月8日

富士通株式会社

先端コンピューティング開発本部

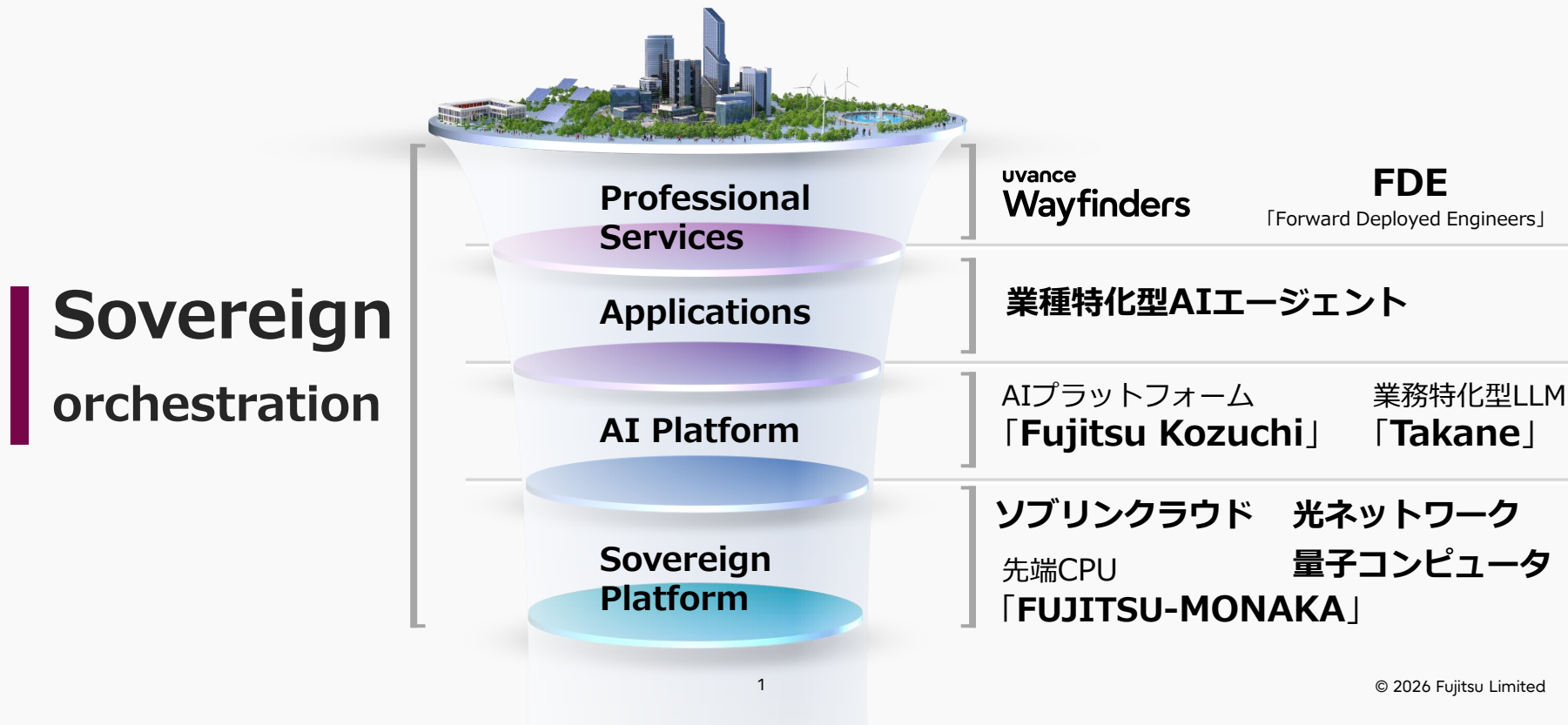
ソフトウェア開発統括部

平井 浩一

この成果は、NEDO（国立研究開発法人新エネルギー・産業技術総合開発機構）の補助事業（JPNP21029）の結果得られたものです

# 富士通が目指す信頼できるテクノロジーの提供 FUJITSU

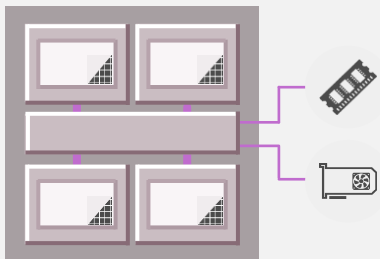
独自技術でCPUからAIプラットフォームまでを構築、実践知をもってお客様に実装



# 富士通プロセッサロードマップ

常に世界一を目指し、日本のHPCを支えてきた





**Armv9-A Architecture**



**3D chiplet**

- Core die 2nm
- SRAM die/IO die 5nm



**Ultra low voltage for energy-efficiency**



**DDR5 12 channels**



**Air-cooled, Water-cooled**



**Arm SVE2 for AI and HPC**



**144 cores x 2 sockets (288 cores per node)**



**Confidential Computing for security**



**PCI Express 6.0 (CXL3.0)**

この成果は、NEDO (国立研究開発法人新エネルギー・産業技術総合開発機構) の助成事業の結果得られたものです。

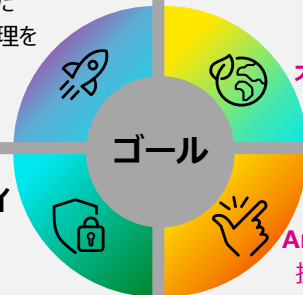
## デジタル社会を実現する 高性能・省電力・国産プロセッサ

**高速なデータ処理基盤**

AIワークロードを中心とした  
コンピューティングの高速処理を  
実現(競合CPU比2倍)

**省電力とパフォーマンスの両立**

高い電力効率により  
CO2排出と電力コストを  
大きく削減(競合CPU比2倍)



**信頼性とセキュリティ**

メインフレームで培ってきた  
安定稼働技術と  
クラウド活用に向けた  
CCにより高いセキュリティを実現

**使いやすさ**

Armソフトウェアエコシステムを  
拡大し、性能と利便性を両立  
サービス・ソフトウェア・ハードウェアの  
全体を通じた協調設計(Co-design)

自社設計のマイクロアーキテクチャや  
低電圧技術などの当社独自技術により実現

**FUJITSU-MONAKAの  
戦略的技術連携**



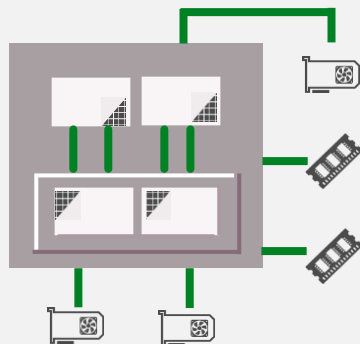
**Supermicro社**



**AMD社**



**NVIDIA社**



**Armv9-A Architecture**



**2nd Gen 3D chiplet**

- Core die **1.4nm**
- SRAM die/IO die TBD



**NVLink-Fusion(TM) for GPU connectivity**



**Ultra low voltage for energy-efficiency**



**Air-cooled, Water-cooled**



**Arm SVE2 + SME2 for DC, AI-HPC**



**> 200 Cores x 2 sockets (> 400 cores per node)**



**Confidential Computing for security**



**NPU integration (Optional)**



**PCI Express 7.0**

## デジタル社会を実現する 次世代高性能・省電力・国産プロセッサ

### 高速なAI処理基盤

AIワークロードを中心とした  
コンピューティングの  
高速処理を実現  
(競合CPU比2倍)

**CPU-GPU密結合**

### 省電力とパフォーマンスの両立

高い電力効率によりCO2排出量  
と電力コストを大きく削減  
(競合CPU比倍)

**NPU結合でAI推論の効率化**

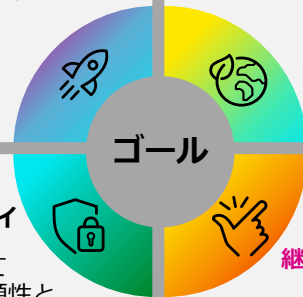
### 信頼性とセキュリティ

メインフレームで培って  
きた安定稼働技術、信頼性と  
ソブリン性を担保し、CC技術で  
堅牢なセキュリティ基盤を提供

### 使いやすさ

**継続的なエコシステムの拡大**

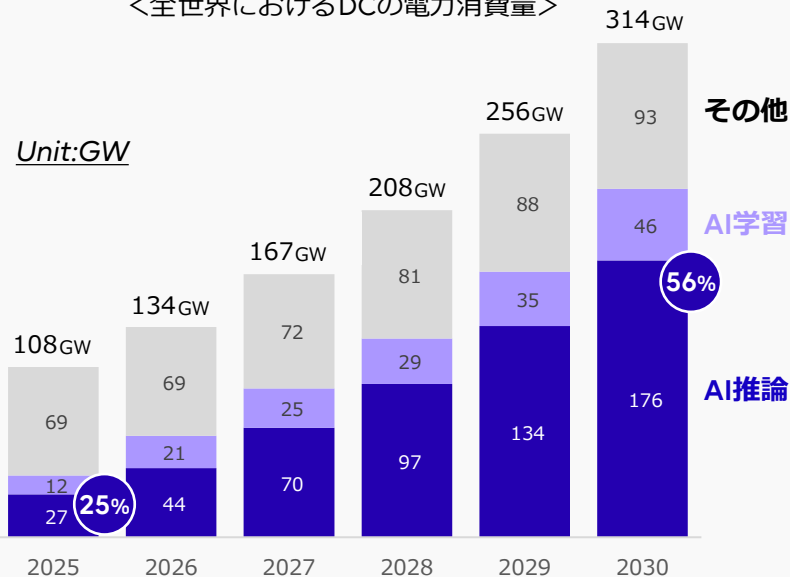
富岳、MONAKAで培ってきた  
ソフトウェア資産を利用可能



# 1 AIを支えるMONAKAシリーズ

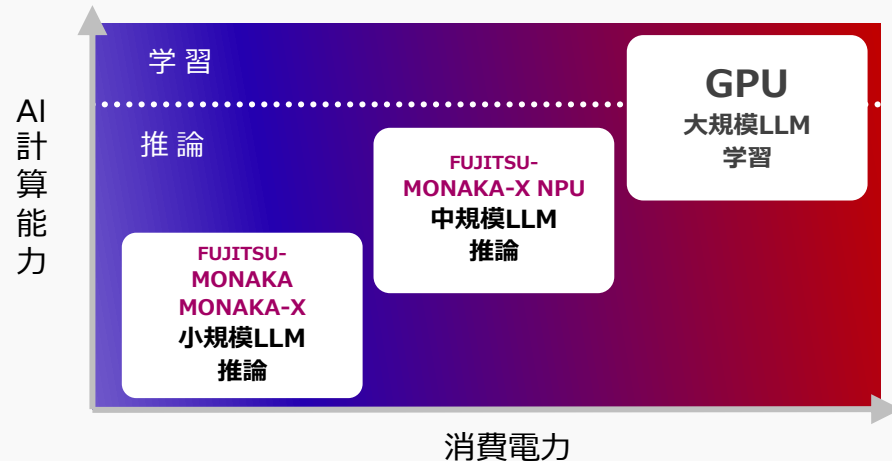
## 今後5年でAI推論需要が拡大

＜全世界におけるDCの電力消費量＞



出典:Omdia AI power capacity Summary 20251203

## 全てのAIワークロードで ベストな選択肢を提供



## 2 エンタープライズ向け生成AI「Takane」

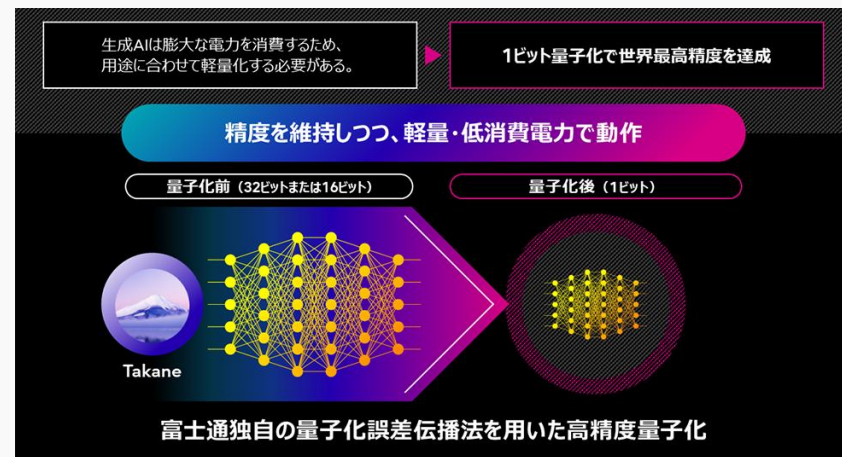
- AIの軽量化・省電力を実現する強化された富士通の生成言語モデル

大規模モデルに匹敵する十分な能力

Takane-2.0-32B  
ガバメントAI 源内における試用開始

Sovereign環境に向けて  
ローカルLLMによるCoding能力強化

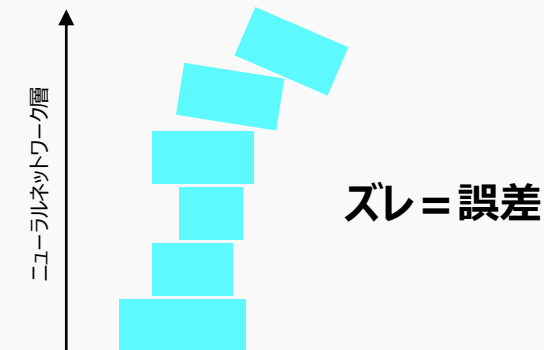
1ビット量子化技術により  
89%精度維持率、94%軽量化



# ◆富士通の1ビット量子化技術

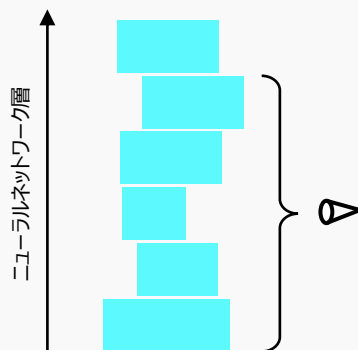
量子化誤差伝播法を用いた高精度量子化により、世界最高の精度を達成

## 従来技術



各層にズレが発生し、層数が増えるにつれズレが大きくなり崩れる（精度崩壊）

## QEP ( Quantization Error Propagation )



### 誤差伝搬法

ズレを見ながら、全体がまっすぐ積み上がるように調整する方法

### キャリブレーションデータ活用

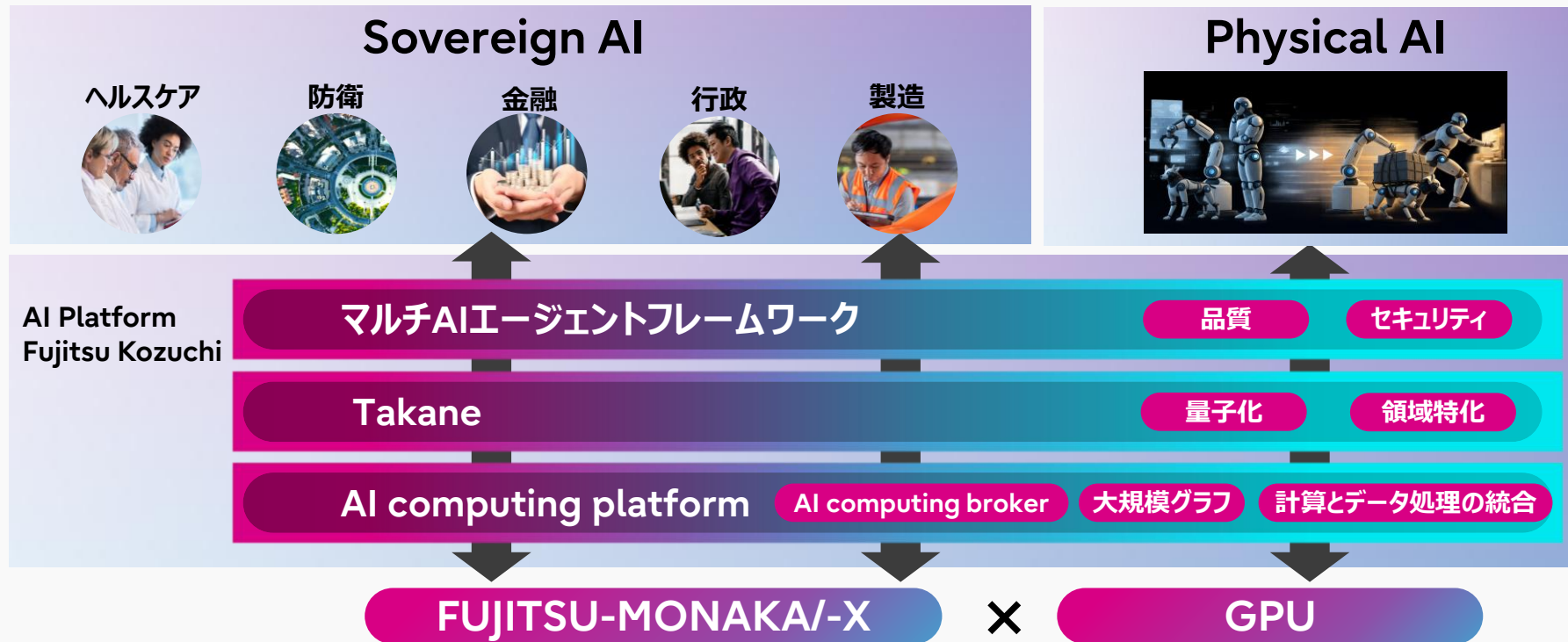
サンプルデータを用いて、安定する位置をあらかじめ決めておく

## QQA (Quasi-Quantum Annealing)

デジタルアニーラで培った世界最高性能の大規模最適化アルゴリズム

### 3 自立型AIプラットフォーム「Fujitsu Kozuchi」

- コンピューティングとの有機的垂直統合で現場で使えるSovereign/Physical AIを提供



# 3 自立型AIプラットフォーム「Fujitsu Kozuchi」

## ● ソブリン性の高いAIを実現する手段としてオンプレミスでも活用可能

### 1. 専有環境を実現する インフラ基盤



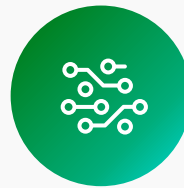
- オンプレミスに対応し、データを外部に出さないAI活用を実現
- 導入から運用高度化までを一貫して支援
- PRIMERGYベースの高性能GPU基盤を採用\*

### 2. 生成AIトラスト技術



- 9,000種超の脆弱性に対応したLLM脆弱性スキャナー
- ガードレールによる不正入力・不適切出力の抑止
- 非専門家でも安全性を確保できる自動運用

### 3. 高精度生成AIモデルと カスタマイズ・軽量化技術



- 高精度日本語LLM「Takane」を中核に採用
- 内製型ファインチューニングで業務特化型モデルを作成／改善
- LLM量子化技術により低コスト・高効率なAI運用を実現

### 4. AIEージェント開発 効率化技術



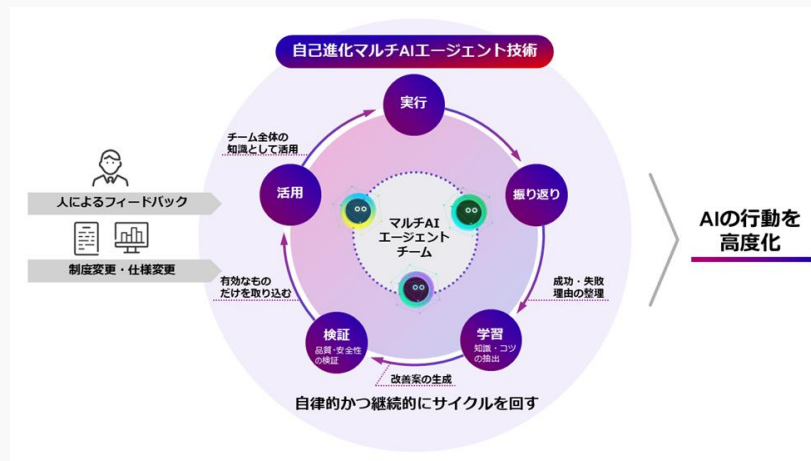
- ローコード／ノーコードで業務AIを迅速に構築
- MCP対応により既存システムと柔軟に連携
- 複数AIEエージェントの協調動作による高度活用

## 4 AI for Scienceにおける取組

様々な実シミュレーションを  
AIサロゲートモデルで高速化

AI Agentによる業務支援

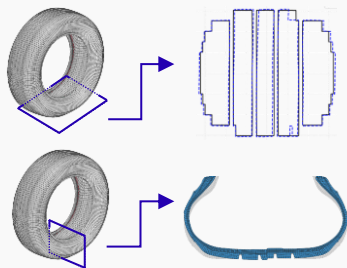
- 1 **タイヤ構造解析サロゲート**  
大規模メッシュへのGNN適用
- 2 **宇宙サロゲート**  
太陽フレアを予測するPINNs開発
- 3 **分子動力学サロゲート**  
構造形成プロセスの解析



4 継続的に安全に進化する  
マルチAIエージェント技術

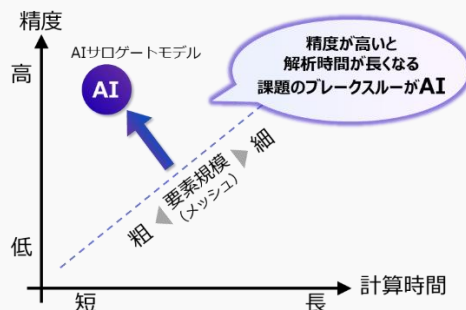
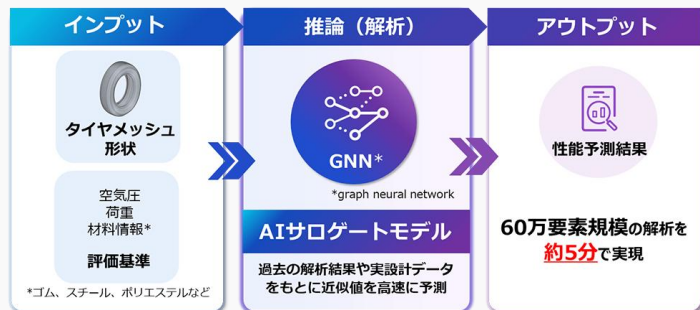
# 4-1 タイヤ構造解析サロゲート

複雑材料かつ大規模なタイヤの路面接地変形挙動をAIにより高速推論



- ✓ AIサロゲートモデルにより**タイヤ構造解析時間を約90%短縮**
- ✓ メッシュ構造をグラフベースで学習できるGNN技術を活用し、**大規模・複雑なタイヤ構造の変形挙動を高精度に再現**
- ✓ 設計DXを加速し、**リアルタイムな設計判断を可能に**

➡ 2026年12月までにMONAKA検証機での実証開始予定



2026年6月3日 両社共同プレスリリース配信

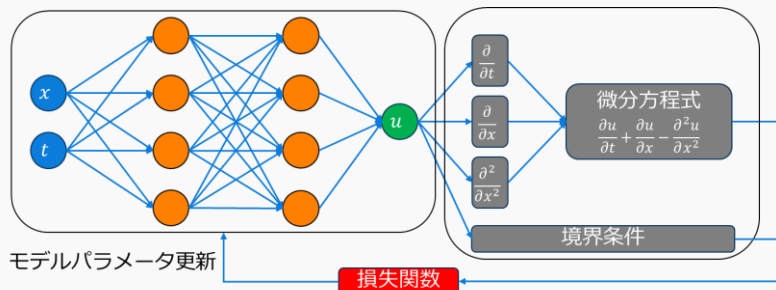
- ・DUNLOP [https://www.srigroup.co.jp/newsrelease/2026/sri/2026\\_055.html](https://www.srigroup.co.jp/newsrelease/2026/sri/2026_055.html)
- ・富士通 <https://global.fujitsu/ja-jp/pr/news/2026/06/03-01>

## 4-2 宇宙サロゲート

### PiNNsを用いたサロゲートモデル

#### PINNs : Physics-Informed Neural Networks

- 微分方程式を満たすように訓練を行うニューラルネットワーク
- 微分方程式の独立変数を入力とし、その独立変数の値における解の値を出力
- 微分方程式/境界条件のそれぞれを満たすほど0に近づく損失関数を定義
- 損失関数が0となるようにモデルパラメータを繰り返し更新

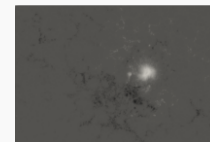


### 太陽フレアの発生予測

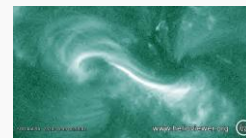
発生予兆であるコロナ層の磁力線のねじれを検出するため、磁場の外挿処理をPINNsで実装

適用事例：AR13848（X1.8フレア発生）

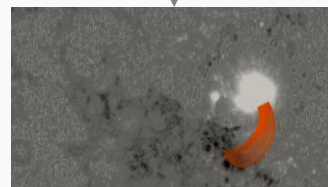
光球面の磁場



EUV観測結果



従来手法  
: MHD緩和法



PINNs

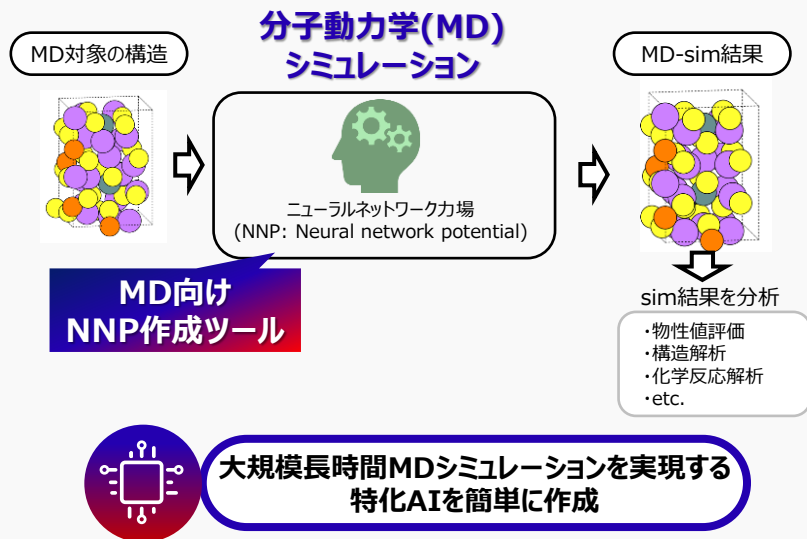


日本天文学会2026年春季年会発表

<https://www.asj.or.jp/nenkai/archive/2026a/pdf/M22a.pdf>

# 4-3 分子動力学サロゲート

## ターゲット材料の化学反応のシミュレート

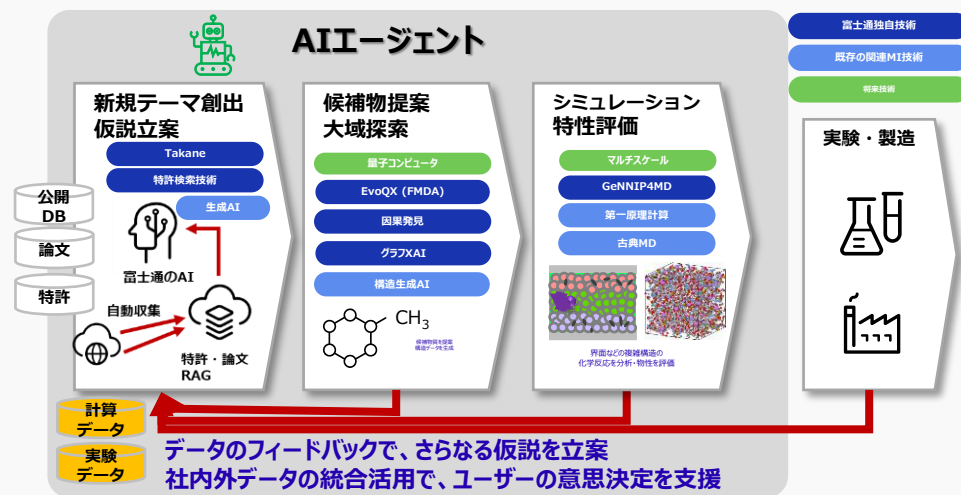


発表論文

<https://arxiv.org/a2411.17191bs/>

<https://arxiv.org/abs/2503.20412>

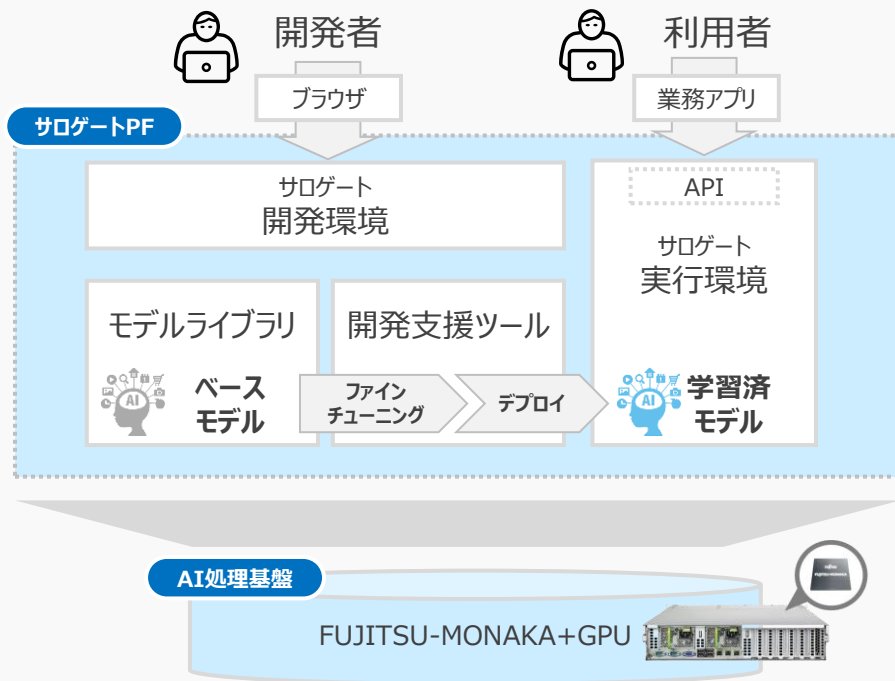
## AIエージェントを用いて自動化



# ◆ サロゲートプラットフォーム powered by MONAKA

FUJITSU

サロゲートモデルの開発を環境最適化からモデルのデプロイまで  
一貫支援するオールインワンプラットフォーム



## 開発 実行環境



- ・インタラクティブに実行可能な開発環境
- ・ノーコード化 (Webアプリ化)

## モデル ライブラリ



- ・実用的な学習済みベースモデルを複数実装予定
  - 風洞シミュレーション向け汎用モデル
  - 車内空調シミュレーション向け汎用モデル

## 開発支援 ツール



- ・チューニングの自動化
- ・ワークフローパイプライン化
- ・検証履歴やコード、モデルの一元管理

## AI処理基盤 (MONAKA)



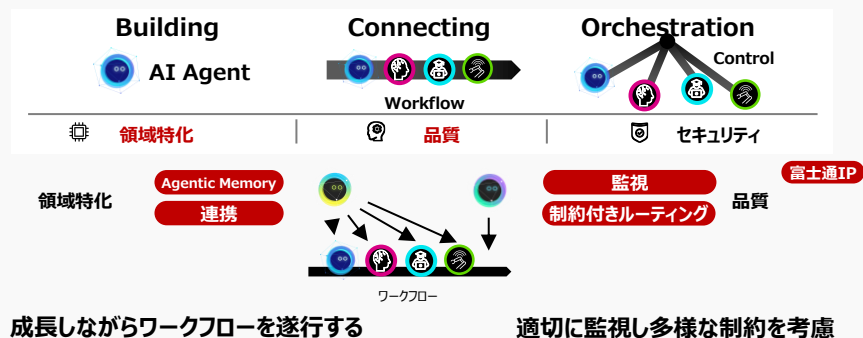
- ・パフォーマンスと電力効率の最適化
- ・セキュアで柔軟なソブリン環境
- ・学習環境に最適化されたAIライブラリ

Fujitsu Research Portal : AIサロゲートモデル検証プラットフォーム

<https://documents.research.global.fujitsu.com/ai-surrogate-model-trial-platform/>

# 4-4 自己進化マルチAIエージェント技術

## Multi AI Agent Framework



AIエージェントを領域特化、  
高品質でセキュアなワークフロー自動化

## 業務特化型LLM「Takane」の 自動的な強化

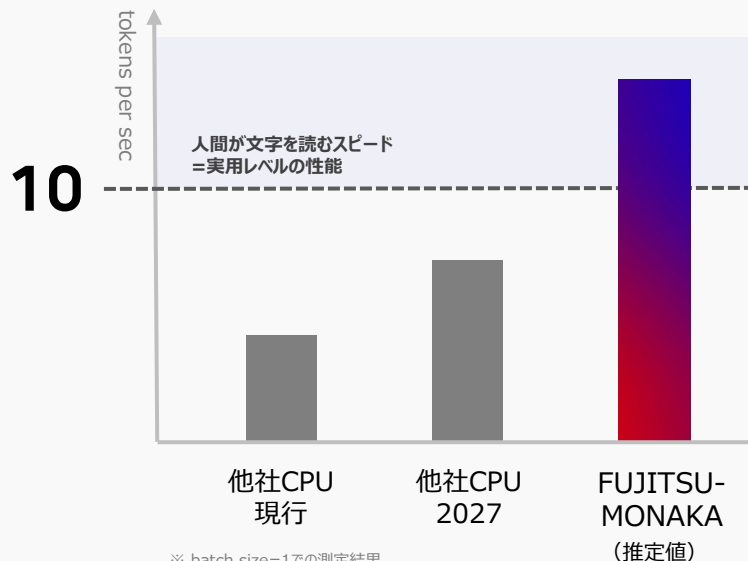


マルチAIエージェントが自律的に実行・最適化

# ◆ 高精度LLM推論へのFUJITSU-MONAKA適用

FUJITSU-MONAKA+ソフトウェア最適化で70B規模の推論性能を実現

## Llama-3.3 性能比較 (70B)

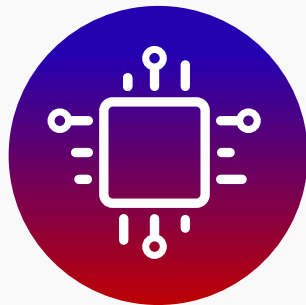


## ソフトウェア最適化

1. SIMD機能活用
2. スケーラビリティ改善
3. 高精度1bit/2bit量子化活用
4. i8mm行列乗算の拡張命令活用

# 8 今後もHPC×AIで進化を続ける

FUJITSU-MONAKA×AI×ソフトウェアの力でAI for Scienceを実現



## 1 計算基盤の継続的な開発

省電力・高性能な  
HPC/AI計算基盤として  
FUJITSU-MONAKAシリーズ  
を継続的に開発・提供



## 2 新たなAI技術の開発

サロゲートモデル、  
マルチAIエージェント、  
業務特化型LLM など  
新AI技術を開発・提供



## 3 エコシステムの確立・拡張

AI for Scienceの民主化に向け、  
AI/HPCに関するソフトウェア  
エコシステムを確立・拡張を推進

Thank you