

# 研究戦略説明会

Fujitsu

Research Strategy

Briefing Session

## AI領域の研究戦略

エンタープライズ

生成AIフレームワーク

2024年6月4日

富士通株式会社

富士通研究所人工知能研究所所長

**園田 俊浩**

大規模



言語に加えて、映像や音声にも対応  
クラウド上で広範囲に公開

GPTに代表される、大規模汎用利用の

## 大規模汎用モデル - LLM

100B

### 企業業務を支援する特化型

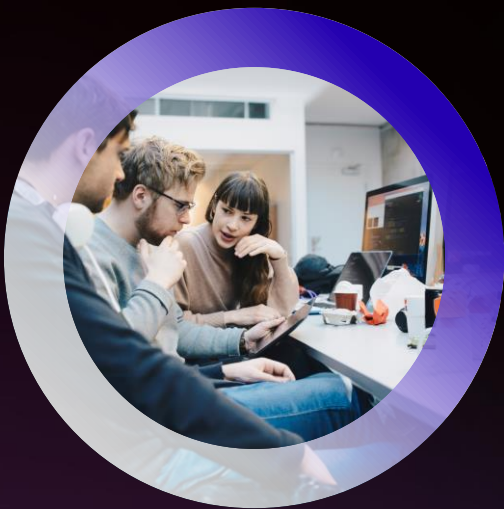
・カスタマイズ性 ・企業規則への準拠



サイズをコンパクトで効率的に最適化した

## 小中規模言語モデル - SLM

小・中規模



汎用的な大規模モデルと小中規模の特化型モデルの市場が拡大  
➤ 企業ニーズを満たす特化型モデルに注力

## 生成AIの企業ユースにおける3つの課題

1. 企業が保有する多様で大量なデータを扱えない
2. 業務ノウハウやプロセスに特化したモデルを迅速に生成できない
3. 企業における規則や法令に準拠させることが困難

企業における生成AIの活用上の課題を解決し、セキュリティ不安を払しょくする

## エンタープライズ生成AIフレームワーク

企業における生成AI活用を牽引するグローバルトッププレイヤーを目指す

グローバル12万4千人の従業員が  
生成AIを活用できる環境を整備し  
社内実践

「Fujitsu Kozuchi」上で  
企業向け対話型生成AIを公開

「富岳」で学習した大規模言語モデル「Fugaku-LLM」を開発

- 130億パラメータのモデルを一から独自データで学習
- 富士通は、演算と通信の高速化、事前学習と学習後のファインチューニングを担当

# 富士通のエンタープライズ生成AIフレームワーク

FUJITSU

企業における生成AIの活用上の課題を解決、セキュリティ不安を払拭する  
富士通のエンタープライズ生成AIフレームワークを構成する3つの技術

世界最高  
性能達成

ナレッジグラフ  
拡張RAG

多様で大規模な  
企業データに対応

1000万  
文字対応

世界最高  
レベル性能

生成AI混合技術

変化する企業ニーズに  
柔軟に対応

0  
カスタマイズ

世界初

生成AI監査技術

挙動制御で  
AI活用の不安払拭

法令遵守

# エンタープライズ生成AIフレームワークの全体像

## ステップ 2 ランタイム

クエリーに企業が保有するデータを紐づける

特化型生成AIを選択  
なければ組合せて  
AIモデルを構築

生成AIが法令や規則に準拠しているかチェック

## ステップ 1 準備

多様で大規模な企業データからナレッジグラフを準備

製品マニュアルや設計書

監視ビデオ

ネットワークログ

法令や企業規則

クエリーの送信

クエリー指向ナレッジ

特化型生成AI

入出力制御・説明付与

ナレッジグラフ  
拡張RAG

生成AI  
混合技術

生成AI  
監査技術

ナレッジグラフ

生成AIや既存AI

ナレッジグラフ

RAG: Retrieval Augmented Generation(検索拡張生成) LLMにて学習外の知識を取得して利用する仕組み / ナレッジグラフ: さまざまな知識を体系的にグラフ構造で表した知識のネットワーク



# ナレッジグラフ拡張RAG

## 1,000万文字以上のドキュメント全体を俯瞰して高精度に分析

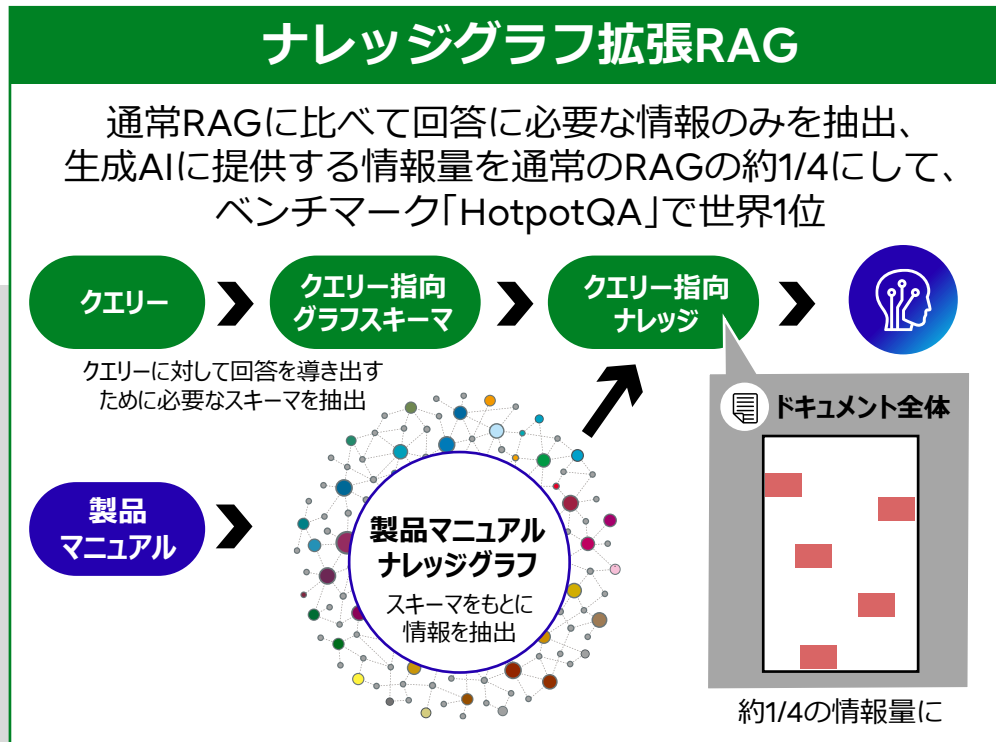
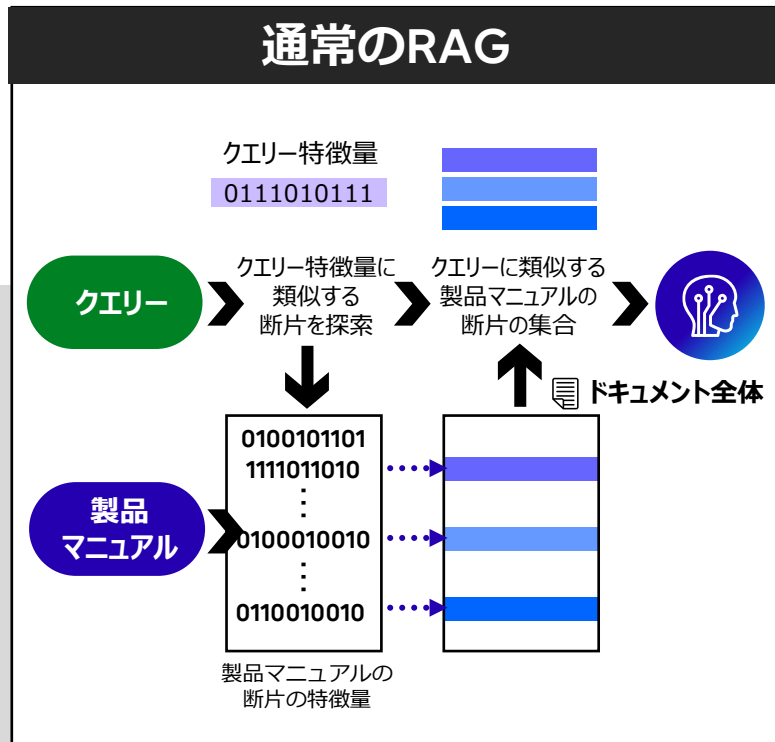
- 製品マニュアルなどの企業が取り扱う大規模データを逐次的に処理し、ナレッジグラフを生成し、効率的に大規模データを処理
- クエリーに応じてナレッジグラフから必要情報抽出、生成AI推論補助



実績

複雑な質問応答の精度を測定するベンチマーク「HotpotQA」で世界1位を獲得

## 通常のRAGに対して高精度な応答を生成可能





# ナレッジグラフ拡張RAGの適用事例

DEMO展示

1000万  
文字対応

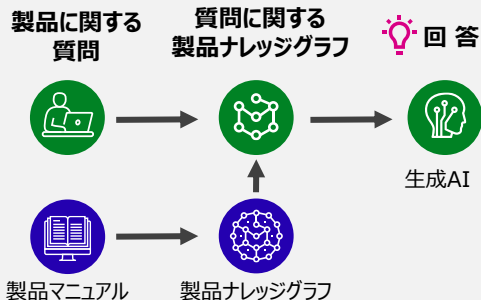
FUJITSU

## 製品マニュアルQ&A

製品マニュアルは1,000万文字以上  
全体を俯瞰した回答ができない

1,000万文字の製品マニュアル  
Q&Aで効果確認

複数ページにまたがる情報を  
適切に統合して最適な回答を生成

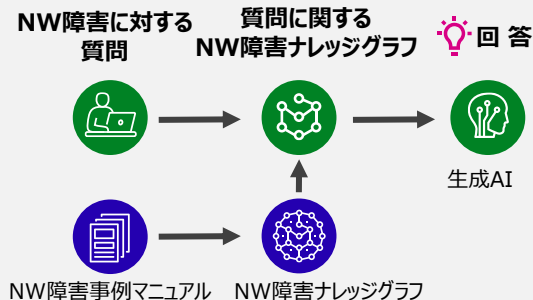


## ネットワークログ解析

大量なネットワークログや過去の障害  
事例からNW障害の原因を発見できない

モバイルネットワーク  
接続障害適用、効果確認

117の異なる障害事例から  
ナレッジグラフを生成、原因候補を  
網羅的に列挙し障害復旧を効率化

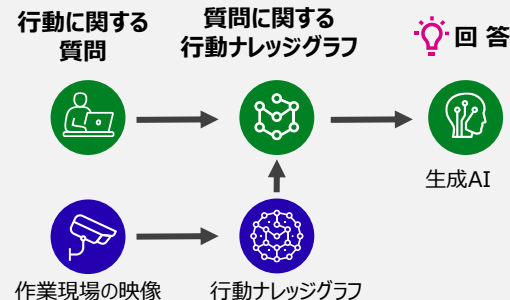


## 映像による作業分析

長期間にわたる大量な映像データの  
集計や統計情報の取り扱いができない

作業現場の映像から長期間の  
作業者の状況を確認可能

実倉庫で検証中





## プロンプトエンジニアリングやファインチューニングなどの顧客によるカスタマイズ不要で、効果が高い特化型生成AIを容易に自動生成

- **特化型生成AIを選択** クエリー特性\*1とモデル特性\*2からタスク実行に必要なAIモデルを選択
- **特化型生成AIの自動生成** 適切なAIモデルがない場合は、必要なAIモデルを自動生成



**実績**

映像検知においてGPT-4Vと同等、日本語オープンモデルで最高性能達成  
契約書順守チェックやサポートディスク効率化といった企業業務に特化したタスクで効果確認

\*1 クエリー特性 ユーザのクエリーの特性を表す指標でこのタスクを処理するモデルを選択するときに利用 \*2 モデル特性 特化型モデルを生成するときに追加される企業情報の特性。ユーザからのクエリーを処理するモデルとしての適合性チェックのために利用

## 契約書順守チェック



ソフトウェア契約書と利用状況の確認に膨大な時間がかかっている

契約内容との突合

業務知識での補完

富士通の確認ノウハウを  
生成AIに教える必要あり

数カ月かけてプロンプトで  
業務知識を扱えるようにすることなく  
特化型生成AIを自動生成

30% 工数削減



契約書  
利用状況

契約書  
分析AI

利用方法  
確認AI

## サポートデスクの効率化



インシデント対応担当者の配置が俗人的で遅延が発生

担当者スキル

SLA遵守

緊急度

予測・最適化モデル開発に  
専門家が必要

複雑なタスクアサインメントを解く  
特化型生成AIを自動生成

作業効率 25% 向上



インシデント

予測

最適化

担当割振り

## ドライバー最適配置



物流2024年問題  
20万人のドライバー不足

法令遵守

人手不足

最適化の定式化に  
専門家が必要

専門家が必要な定式化を  
即座に実行できる  
特化型生成AIを自動生成

計画作成時間 95% 削減



運送依頼

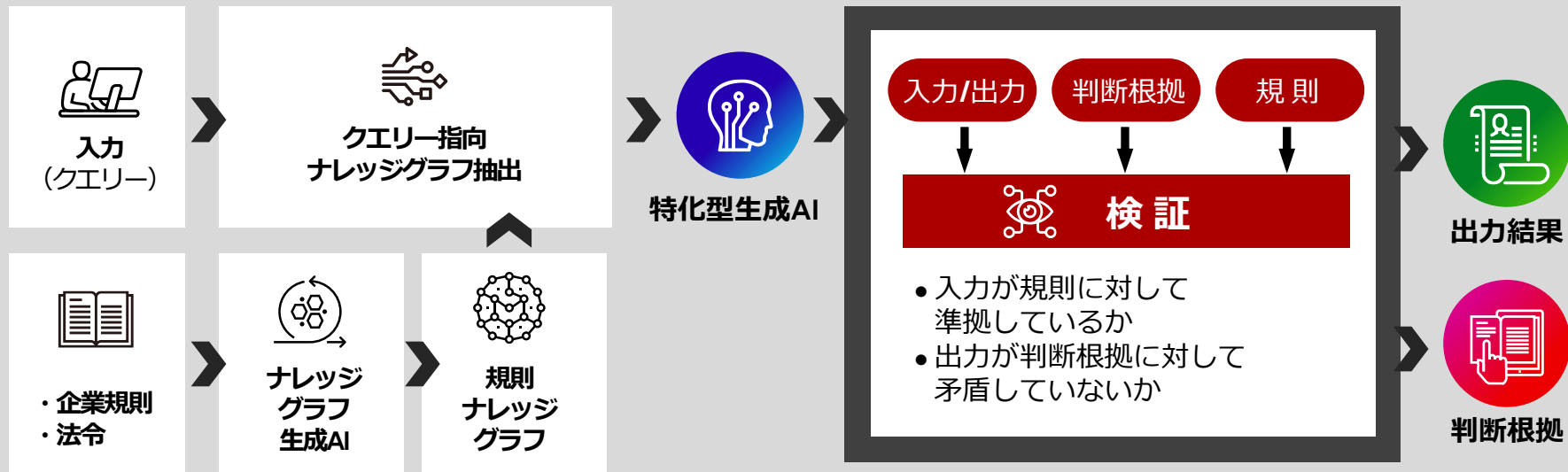
最適化

運送計画



## ナレッジグラフで生成AIの挙動を制御し、企業規則や法令に準拠

- 法令や企業規則に対応するナレッジグラフを活用して、**入力の規則準拠を検証**
- 生成AIが出力を導出した根拠を解析することで、**判断根拠の説明性付与、ハルシネーションを判定**



## 規則準拠検証 と 出力根拠の説明付与



検証

入力



規則



第二条 十一イ；自転車⇒軽車両  
第十七条；車両⇒車道を通行  
第十七条 4；車両⇒左側部分を通行

入力が、規則に対して準拠しているか



入力



入力が規則に準拠しているか判断するよう指示



規則



生成AI



判断

以下の状況が違反しています。

1. 車道を走っている自転車は車道の左側を走らなければいけません。
2. ヘルメットをかぶらずに車道を通行する自転車は、ヘルメットをかぶるように努めなければなりません。

出力

以下の状況が違反しています。

1. 車道を走っている自転車は車道の左側を走らなければいけません。
2. ヘルメットをかぶらずに車道を通行する自転車は、ヘルメットをかぶるように努めなければなりません。

判断根拠



出力が、判断根拠に対して矛盾していないか



生成AIが規則準拠を判断した根拠を解析

出力



判断根拠



矛盾チェック

判断根拠と出力の矛盾をチェック



矛盾なし

ナレッジグラフ  
拡張RAG



## AIサービス「Fujitsu Kozuchi」

エンタープライズ生成AIフレームワークを  
7月より順次公開予定

企業における生成AI活用を牽引する  
グローバルトッププレイヤーを目指す

# 研究戦略説明会

Fujitsu

Research Strategy

Briefing Session

次世代グリーンデータセンター向けプロセッサ

## FUJITSU-MONAKA

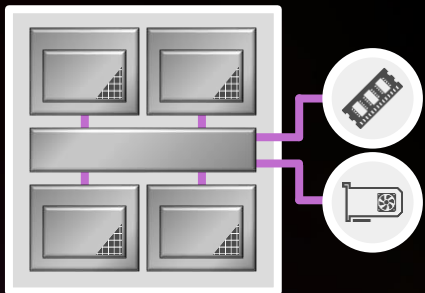
2024年6月4日


富士通株式会社


富士通研究所 先端技術開発本部 本部長

**新庄 直樹**


# FUJITSU-MONAKA




 Armv9-A Architecture


 3D chiplet  
• Core die 2nm  
• SRAM die/IO die 5nm

 Ultra low voltage  
for energy-efficiency

 DDR5 12 channels

 Air cooling

 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)

2027 リリース

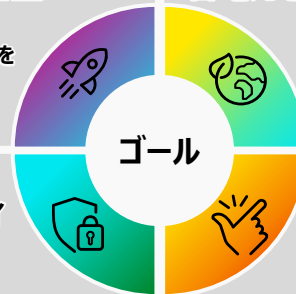
## カーボンニュートラルなデジタル社会を実現する 次世代高性能・省電力・国産プロセッサ

### 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  
利用可能  
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

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



# 様々な用途での活用を目指し、各種分野との共創を開始



Arm-based CPU  
FUJITSU-MONAKA

FUJITSU-MONAKA  
Software



## 高まるデータセンターのAI需要に応える

AI性能の追求

他社CPU比2倍の  
性能と電力効率

AIソフトウェア拡充

幅広い  
ソフトウェアスタック

# AIやHPCを始めとする広い領域のソフトウェアをカバー

<b>Product Delivery</b>	<b>Customer Use Cases</b> <ul style="list-style-type: none"> <li>Surrogate Models SVR</li> <li>LLM Software Applications</li> </ul>	<b>Fujitsu Computing as a Service</b> <ul style="list-style-type: none"> <li>Scikit Learn Use Cases</li> <li>Hugging Face Use Cases</li> </ul>	<b>Fujitsu Kozuchi</b> <ul style="list-style-type: none"> <li>Causal Inference</li> <li>Ambient Authentication</li> </ul>							
<b>Open-Source Contributions</b>	<b>API Microservices Platform (FUJITSU-MONAKA Green HPC API Server) @ FR IPL*</b>									
	<b>OpenMathLib /OpenBLAS</b>	Math Library, NumPy, OpenMP	<b>UXL foundation</b>	oneAPI Ecosystem, oneDAL & oneDNN	<b>PyTorch</b>	FUJITSU-MONAKA ARM Improvements	<b>Linaro</b>	Kubernetes and OpenStack OSS		
<b>Software Delivery</b>	PyPi	Docker	Containers	Reference Implementations	Computing Workload Broker	Continuous Integration and Deployment using MOCHI and Konark Platform				
<b>Collaborations</b>	<b>Internal Teams</b>					<b>External Organizations</b>				
	SW R&D	Math Lib	Platform	AI Solutions	Computing	Compiler	ARM	UXL	MIT	IISc
<b>AI Software Frameworks</b>	Scikit-Learn Multithreading XGBoost NumPy Pandas BLAS <b>Machine Learning</b>		LLM's Vision NLP Hugging Face TensorFlow/PyTorch OpenVINO oneDNN Inductor <b>Deep Learning</b>		PostgreSQL PySpark VectorDB Data Intelligence <b>Big Data Analytics</b>		Red Hat Secured HW/SW Software Guard Extensions OpenShift Confidential Computing <b>Data Security</b>			
<b>Software Stack Selection</b>	<b>Quantitative Metrics</b>				<b>Qualitative Metrics</b>					
	Downloads	GitHub	Market Adoption	Search Trends	Arm Enablement	Release Freq.	Innovation Scope	Use Case		
<b>Cutting Edge Applications</b>	<b>Healthcare</b> <ul style="list-style-type: none"> <li>Drug Discovery</li> <li>Gene Prediction</li> </ul>		<b>Manufacturing</b> <ul style="list-style-type: none"> <li>Defect Detection</li> <li>Preventive Maint.</li> </ul>		<b>Retail</b> <ul style="list-style-type: none"> <li>Recommendation</li> <li>SCM Forecasting</li> </ul>		<b>Banking</b> <ul style="list-style-type: none"> <li>HF Trading</li> <li>Fraud Detection</li> </ul>			

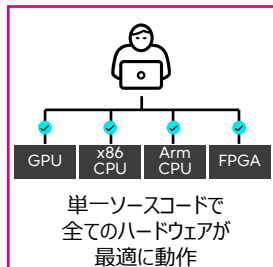
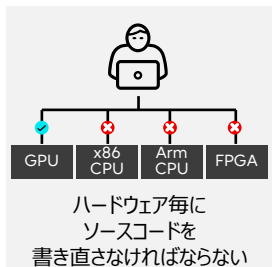
# 導入障壁を下げるソフトウェア技術開発例

## 単一コードで異種AIアクセラレーターを活用するUnified Acceleration技術の開発



- 様々なCPUやアクセラレーターを単一ソースコードで活用可能にするための団体UXL foundationを発足メンバー企業として牽引
- 富士通はArm CPU FUJITSU-MONAKAをAIアクセラレーターとして活用するための基盤ソフトウェアをグローバルに開発する
- 2027年リリースの時点でお客様がFUJITSU-MONAKAのAI性能を手軽に、最大限活用できる環境の構築を目指す

### Unified Accelerationのメリット



高い環境移行コスト

環境移行コストの低減

メンテナンス性の低下

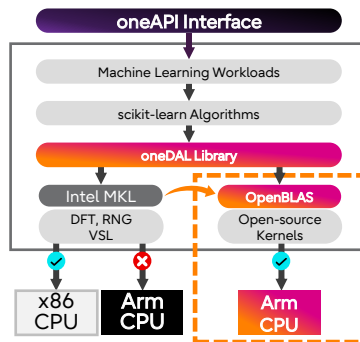
最適なハードウェアが選択可能

FUJITSU-MONAKAのAI活用拡大に繋げる

### 最新事例：

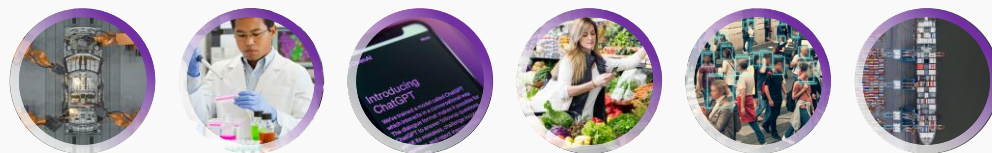
### oneDALのArm enablementに世界で初めて成功

Intel MKL固有の数学関数を  
オープンソースであるOpenBLASの関数に置き換えることに成功



高速に処理する演算ルーチンを手軽に大規模演算で  
使用することが可能に

Arm enablementを拡大し  
AIソリューション開発基盤を  
構築する



Fujitsu Kozuchi

FUJITSU-MONAKA & Tuned AI softwares

Application	AI 音声認識          サロゲートモデル          LLM 活用基盤					
Library Framework	OpenBLAS	NumPy/SciPy	scikit-learn	oneDNN	oneDAL	PyTorch/TensorFlow
Middleware OS	Linux	Kubernetes	OpenStack	GCC/LLVM	Confidential Containers	
Firmware Hardware	Many Core	High-Capacity Memory	Low Power	Support AI instructions		

FUJITSU-MONAKAは  
幅広い領域で活用可能な  
AIインフラ基盤として  
お客様の課題を解決していきます

# 研究戦略説明会

Fujitsu

Research Strategy

Briefing Session

## 富士通の研究戦略

技術領域の融合による  
新しい価値創出

2024年6月4日

富士通株式会社  
執行役員 EVP  
富士通研究所 所長

**岡本 青史**

# 富士通の研究戦略

FUJITSU

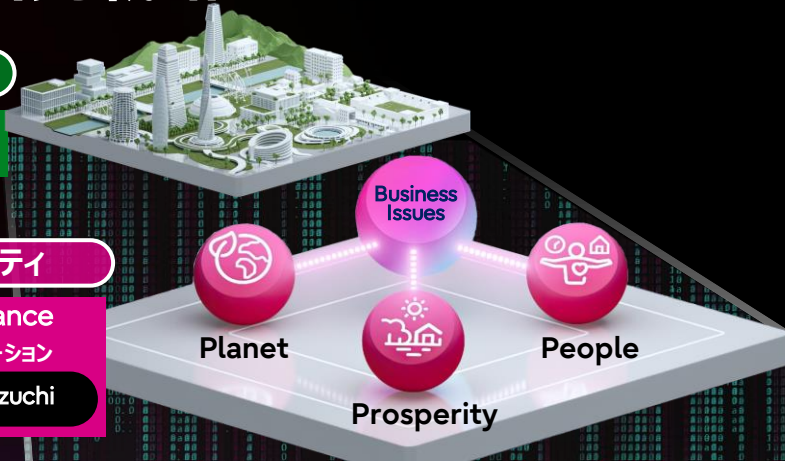
サステナブルな社会

Regenerative enterprise  
(再生型企业) への変革

マテリアリティ

Fujitsu Uvance  
サービスソリューション

Fujitsu Kozuchi



## AIを軸にした 技術領域の融合による 新しい価値創出

Converging technologies

ソーシャルデジタルツインで  
社会課題を解決

Data & Security

偽・誤情報やハルシネーション  
対策技術で安心安全な社会

Computing

AI Computing Brokerによる  
計算コストの大幅削減

量子コンピューターの計算能力による  
圧倒的なイノベーション

Network

大規模・複雑化するネットワークを  
レジリエント化

AI

5 キーテクノロジー



# AI × コンピューティング

FUJITSU

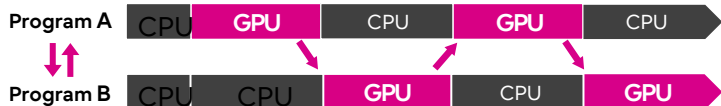
2030年 地球の全電力の **10** %が  
データセンターで消費される

**AIの発展は地球規模の  
電力問題に直結**

データセンターの消費電力を劇的に削減

## GPUを100% フル活用する技術

GPUでモデル計算すべきジョブを  
事前に分析し、動的に割り当てる



様々な効率化方策をとっている  
TSUBAMEでもGPU利用率は30%程度

## GPU削減により 電力量半減へ

AIの計算機リソース縮小で  
年間10TWh以上を削減可能に



日本の約2400万世帯の  
年間電力消費量に相当



# AI × Data & Security

FUJITSU

AIがもたらす偽・誤情報は最大のグローバルリスク  
世界経済フォーラム2024

生成AIや合成コンテンツによる偽情報が、  
選挙プロセスや株式市場操作など  
かつてない社会リスクをもたらしている



# 新たな社会課題への取り組み

## ルールメイキングと偽情報対策技術の開発



国際ガバナンス形成の議論に参加・提言

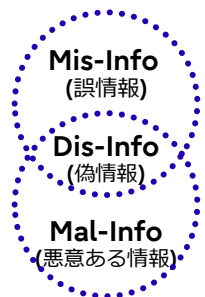
OECD

G7広島AIプロセス

総務省・経産省 AI事業者ガイドライン

## 世界初の真偽判定統合分析システムの構築 新技術

問題となる情報



### 真偽判定統合分析システム



生成元分析

- テキスト
- 画像・動画
- 音声



根拠収集

- 根拠/エンドースメントグラフ

世界初



統合分析

- 矛盾検証
- 根拠分析



判定結果

- 真意判定
- 根拠説明

判定結果と  
文章で根拠を説明

真偽判定

根拠説明



都市は世界のエネルギーの65%以上を消費し  
世界のCO2排出量の70%以上を排出  
欧州委員会2024

AIと人文・社会科学との融合で  
社会問題に取り組む

# 環境, 社会, 経済のトレードオン施策立案

DEMO展示

FUJITSU

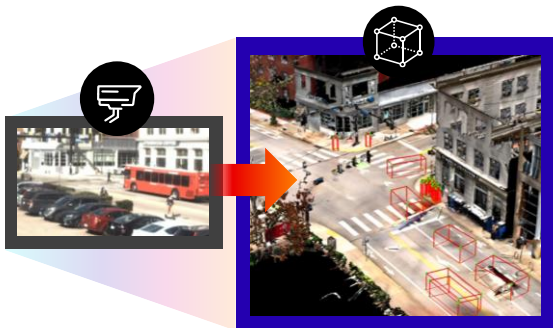
## ソーシャルデジタルツイン

行動選択モデル

合意形成支援

**NEW** リアルタイム3Dツイン生成

**世界初** 1台の単眼カメラでリアルタイムに都市の3Dツインが作れる



モビリティ



エネルギー・環境



防災・防犯



ウェルビーイング



ピッツバーグ

都市交通の安全化



欧州

公共交通機関の利用促進



移動の効率化と脱炭素を両立



インド

EV充電インフラの配置



充電スポットの効率化と脱炭素を両立



山形

地域医療の再編



医療サービス向上とコストを両立



量子がもたらす桁違いの計算機性能で  
AIの世界に革命を起こす



超パーソナライズドAI



大規模マルチエージェントAI



超低消費電力エッジAI

# 世界初の量子機械学習技術

新技術

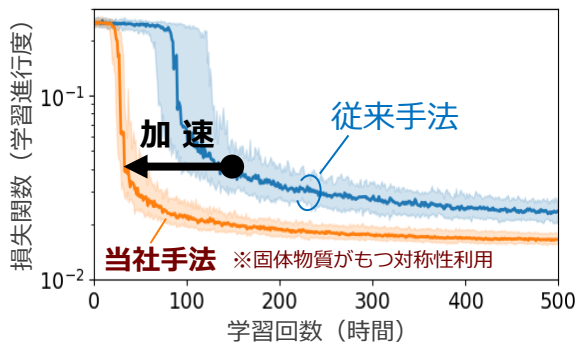
FUJITSU

## ハイブリッド量子プラットフォームの活用開始

### 世界最高速の量子CNN技術

固体物質の性質を予測

磁性体の性質に対する学習

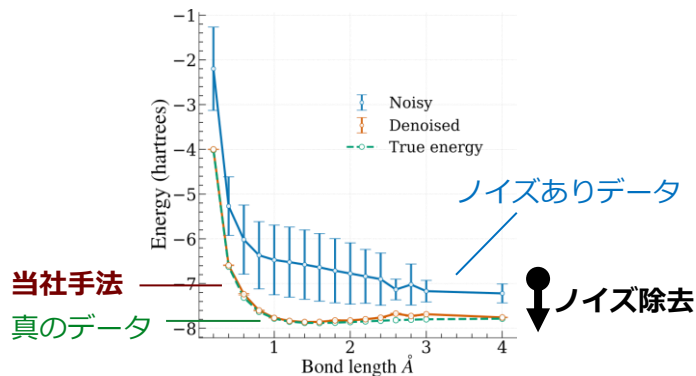


CNN: 畳み込みニューラルネットワーク

### 世界初の量子ノイズ除去技術

量子オートエンコーダを用いてデータ復元に成功

水素化リチウムのエネルギー計算



# AIを軸にした技術領域の融合による新しい価値創出

FUJITSU