

①ポスト5G情報通信システムの開発

(g) 計算可能領域拡大のための計算基盤技術開発

(g1) 量子・スパコンの統合利用技術の開発

[計算可能領域の開拓のための量子・スパコン連携プラットフォームの研究開発](#)

(g2) 超省電力・高密度AI計算基盤技術の開発

[超高効率AI計算基盤の研究開発](#)

(g3) 競争力ある生成AI基盤モデルの開発

[LLMの社会実装に向けた特化型モデルの元となる汎化的LLMに関する研究開発](#)

[自律型エージェントシステム向け高効率基盤モデルの開発](#)

[オープンかつ日本語に強いGPT-3級大規模言語モデルの構築](#)

[厳密さが要求されるビジネス用途におけるハルシネーションを大幅抑止した基盤モデル](#)

[完全自動運転に向けたマルチモーダル基盤モデルの開発](#)

[多様な日本語能力の向上を目指した公開型の基盤モデル開発](#)

[100B/1Tパラメータからなる大規模マルチモーダル基盤モデルの構築](#)

計算可能領域の開拓のための量子・スパコン連携プラットフォームの研究開発

実施者	国立研究開発法人理化学研究所、ソフトバンク株式会社（共同実施）東京大学、大阪大学
概要	量子コンピュータとスーパーコンピュータ（HPC）を連携するための量子・HPC連携システムソフトウェアを研究開発し、これを用いてこれまでのスパコンのみでは困難だった領域の計算を可能とする量子・スパコン連携プラットフォームを構築する。既存のスパコンのみの計算に対し量子・HPC連携アプリの優位性を実証するとともに、この計算プラットフォームで実行される量子・HPC連携ソフトウェアをポスト5G時代のネットワークで提供されるサービスとして展開する技術を開発する。

1. 開発目的

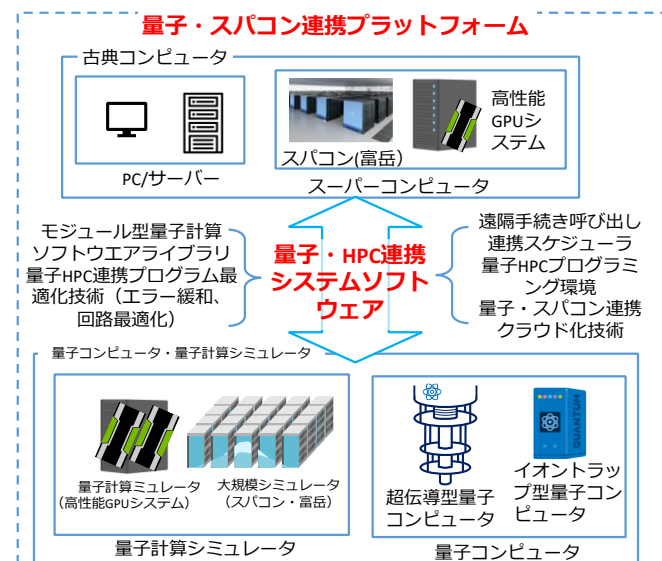
- 量子コンピュータは、従来のコンピュータと全く異なる原理で動作し、情報処理速度の劇的な高速化が期待されるが、現時点では、規模拡大と計算結果の誤り訂正の両立が困難であり、量子コンピュータ単独での実用化には時間を要する見込み。
- 一方で、デジタル化の進展により、情報処理能力の向上が急務であり、量子コンピュータの活用がいち早く求められているところ、古典コンピュータを組み合わせることで活用することが有望視されている。
- 本事業では、世界に先駆けて、量子コンピュータとスパコンを連携利用するためのソフトウェアやプラットフォーム、アプリケーションを開発・構築し、ポスト5G時代で提供されるサービスとして展開する技術としての有効性を実証していく。

2. 開発内容

- 量子・HPC連携ソフトウェア**：スパコンと量子コンピュータを連携させ、最適な計算資源をシームレスかつ効率的に利用するためのシステムソフトウェアを開発。
- モジュール型量子ソフトウェアライブラリ**：アプリ分野に合わせたモジュール型のソフトウェアを整備、量子コンピュータの特性に合わせたエラー緩和処理、回路最適化処理を実現する上位ソフトウェアライブラリを開発。モジュールとして組み合わせることで高度な量子アプリケーションを開発可能とする。
- 量子・スパコン連携PFのクラウド化技術**：事業展開を見据えて、量子アプリケーションの利用を支援するクラウド基盤ソフトウェアを開発。

3. 構築する量子・スパコン連携プラットフォームの構成

- 理研・計算科学研究センター（神戸）及び（和光）に特性の異なる2種類の量子コンピュータを整備。これらと富岳、及び東大・阪大スパコンと連携したプラットフォームを構築。



- ◆ 2026年度に量子・スパコン連携プラットフォームを運用開始し、これを用いて量子・HPC連携アプリケーションの有効性の実証に取り組む
- ◆ 2028年度下期、量子・スパコン連携プラットフォームのプレリリースを計画

超高効率AI計算基盤の研究開発

実施者	株式会社Preferred Networks、株式会社インターネットイニシアティブ、国立大学法人北陸先端科学技術大学院大学
概要	AI技術を活用した計算可能領域の拡大を実現する超高効率AI計算基盤の研究開発を行う。情報通信システムの一部としての超高効率AI計算基盤に向けて、(1) 高い電力効率を実現する次世代AIアクセラレータ・システムおよびその制御技術の研究開発、(2) 大規模商用サービス高密度データセンター基盤技術の研究開発、(3) AI計算基盤の共同利用における実AIワークロードの効率化に関する研究に取り組む。本研究開発では、開発した技術を組み入れたテストベッドを構築し、実AIワークロードを用いた実証実験により、AI計算基盤としての効率化およびその評価を行う。

【超高効率AIアクセラレータ・システムおよびその制御技術の研究開発】

低電圧化 + ソフトウェア技術（電力制御・コンパイラ等） + 高密度実装システム（水冷高密度サーバ）

プロセス微細化（5nm以下予定）

他社GPUの予測値に対して**圧倒的に優位な電力効率性能**を実現

5年後のAIシステム需要を同規模の電力で充足可能な電力効率

2023: PFN MN-Core™ 2 (7nm) 2026: 次世代超高電力効率AIアクセラレータ

- AIに最適化した基本アーキテクチャ
- アプリケーション特性に合わせた各種回路への電力最適分配
- システム全体の最適化制御ソフトウェア

⇒ **AIアクセラレータ・チップおよびノード・ラックレベルでの高効率システムの実現**

【大規模商用サービス展開に向けた高密度データセンター基盤技術の研究開発】

AI計算基盤に求められる**密度**を **直接水冷技術** の応用により実現

- 商用データセンターにおける直接水冷技術利用の**レファレンスモデルと実装**
 - 商用データセンターでの運用に耐える**モジュール化技術**
 - 水冷・空冷ハイブリッド、負荷の差異・変動に対応した水冷仕様等
 - 経済性、継続性、即応性、運用可能性、省エネ性の5要件(*)を充足
- 商用データセンター+AI計算基盤における**省エネ指標**

* 次世代商用データセンターに求められる5要件

- 経済性：IT機器とファシリティ（建物、電気・空調設備）などの**ライフサイクルの異なる機器**の経済的な更改
- 継続性：部品・モジュールの高い**可用性** / 産業としての**継続的に運用可能な事業構造**
- 即応性：需要に応じて**タイムリーに計算キャパシティを増強** / 次世代機器の速やかな導入（新陳代謝促進）
- 運用可能性：IT機器の**負荷変動と消費電力**への**冷却方式**での対応 / 騒音等**労働環境**の低下の防止
- 省エネ性：**データセンター全体**の効率を計測可能な**電力利用効率**の指標

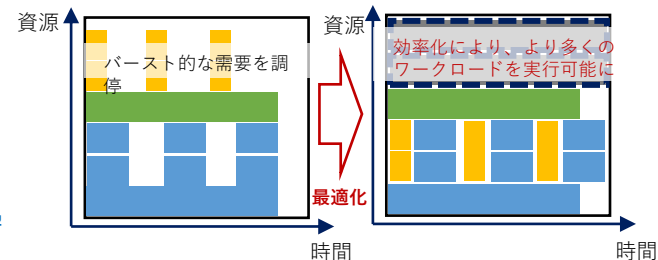
⇒ **AI計算基盤を効率良く収容し、運用可能なデータセンターの実現**

【超高密度AI計算基盤の共同利用における実AIワークロードの効率化に関する研究】

実AIワークロードとAI計算基盤の協調制御による資源割り当て最適化・効率化

→ 上記次世代AIアクセラレータと高密度データセンター基盤技術の成果と合わせて**圧倒的な高実効スループット**（単位時間）の実現を目指す。

⇒ **ハードウェア・ソフトウェア協調制御によるAI計算基盤全体での最適化と効率化の実現**



LLMの社会実装に向けた特化型モデルの元となる汎化的LLMに関する研究開発

実施者 株式会社ABEJA

概要

特化型の元となる汎化的大規模言語モデル（LLM）を研究開発

- ・ オープンソースのLLM（100B相当）をベンチマークとし、評価でトップスコアを達成
- ・ 周辺技術（RAG、Agent）の精度を向上し、データ活用を推進

社会実装に向け、自社ビジネスに関連させた展開を図り、一部モデル・ノウハウ等の成果物も公開・提供

- ・ 研究開発したLLMと周辺技術（RAG、Agent）を、現在提供しているサービスと合わせて広く提供
- ・ 研究開発で得られた成果物(ソースコード・モデル・開発ノウハウ)を公開

実施内容

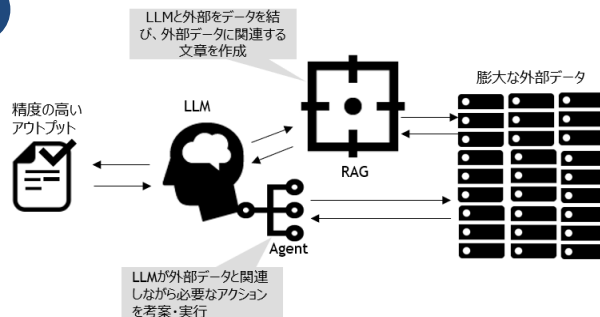
開発される基盤モデル

社会実装の方法

- 1 学習・評価データセット構築
- 2 LLM研究開発
- 3 ベンチマークでの評価実施
- 4 RAG、Agent含む基盤モデルのポストトレーニングの研究開発
- 5 モデルの公開
- 6 ノウハウの公開
- 7 社外コミュニティへの貢献

汎化的な活用を見据えた日本語LLMを研究開発

- ・ オープンソースの既存LLMに対しトップスコアの評価達成
- ・ 周辺技術（RAG、Agent）による精度向上の実現



自社ビジネスとしての展開と成果物の公開

- ・ 現在提供しているサービスと合わせて広く提供
- ・ 研究開発で得られた成果物を公開し、社外コミュニティへの貢献

成果物の公開



ソースコード

汎用モデル、RAG/Agentなどの特化モデルのソースコード



モデル



開発ノウハウ

自律型エージェントシステム向け高効率基盤モデルの開発

実施者	Sakana AI株式会社
概要	<p>大きな計算リソースを必要とする基盤モデルを高効率化するための、重要な技術(蒸留、MoE、強化学習)を用いたモデルを開発</p> <ul style="list-style-type: none"> 高性能・低コストな小規模モデルの開発や、小規模モデルの連携に注力 <p>社会実装に向け、国内でのビジネス展開に加え、一部モデル・ノウハウ等の成果物も公開・提供</p> <ul style="list-style-type: none"> 国内主軸に、各産業分野ごとに特化した製品を開発・販売 開発で得られた成果物(ソースコード・モデル・開発ノウハウ)を公開

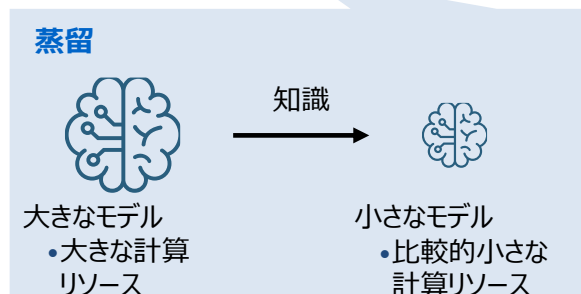
実施内容

- ① 蒸留
- ② MoE(Mixture of experts)
- ③ 強化学習

開発される基盤モデル

大きな計算リソースを必要とする基盤モデルを高効率化するための、重要な技術(蒸留、MoE、強化学習)を用いたモデルを開発

- 高性能・低コストな小規模モデルの開発や、小規模モデルの連携に注力



社会実装の方法

自社ビジネスとしての展開

- パートナー企業と共同で、本事業で開発したモデルを基盤に、各産業分野ごとに特化した製品を開発・販売（国内主軸）

成果物の公開



ソースコード
モデル利用のためのコードのみ



モデル



開発ノウハウ

オープンかつ日本語に強いGPT-3級大規模言語モデルの構築

実施者	大学共同利用機関法人情報・システム研究機構（国立情報学研究所）
概要	<ul style="list-style-type: none"> 日本語に強いGPT-3級(1750億パラメータ)の大規模言語モデル(LLM)を構築 構築したモデル・ソースコード・開発ノウハウ等(議論の過程・失敗も含)を商用利用可能な形で公開 企業・大学との勉強会等で知見を共有、日本の開発力底上げに寄与

実施内容

- ① データ整備**：事前学習用コーパス、評価・チューニングデータ等はLLM-jp*の活動で整備
* LLM-jp = LLM勉強会 <https://llm-jp.nii.ac.jp/>
 - ② 計算環境**：本プログラムから提供される計算資源をフルに使用
 - ③ モデル構築**
 - 1.35Tトークン(日本語約50%)で175Bモデルを事前学習
 - 約20万件のインストラクションデータで追加学習
- ➡ 大規模、かつ高い日本語性能を持つ言語モデルを構築・公開

社会実装の方法

成果物の公開(商用利用可)



ソースコード
前処理・チューニング用ソースコード等

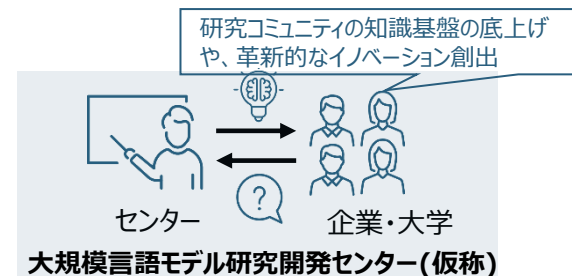


モデル



開発ノウハウ
議論の過程・失敗も含む

国立情報学研究所に「大規模言語モデル研究開発センター」(仮称)を新設し、LLM-jpを発展させた勉強会で上記知見を共有



厳密さが要求されるビジネス用途におけるハルシネーションを大幅抑止した基盤モデル

実施者 スtockマーク株式会社

概要

実ビジネス活用において懸念・障害となるハルシネーション(もっともらしい嘘)を抑制した基盤モデルを開発
 社会実装に向け、国内でのビジネス展開に加え、一部モデル・ノウハウ等の成果物も公開・提供

- 国内で、情報収集・検索SaaSの既存サービスへの組み込み・販売や、PaaS/API化し他社へ販売
- 開発で得られた成果物(モデル・開発ノウハウ)を公開

実施内容

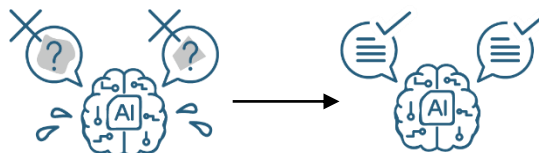
- 1 事前学習・評価
- 2 ビジネス領域に力点が置かれた学習データセットの追加準備・評価
- 3 ②のデータを用いた事前学習・評価
- 4 時事話題等の追加事前学習・評価
- 5 Instruction Tuning・サービスへの組み込み・API化



開発される基盤モデル

実ビジネスでの活用において懸念・障害となるハルシネーションを抑制した基盤モデルを開発

- ChatGPTが正答率40%であるビジネスドメインの質問セットに対して、正答率90%以上
- JGLUEにおいてChatGPT (gpt3.5-turbo) と同水準のスコアを達成



社会実装の方法

自社ビジネスとしての展開

- 自社既存サービスへ組み込み・販売(当面は国内)
 - AI型の情報収集/検索SaaSサービス
 - 大手企業300社に導入済
- PaaS/API化によって他社サービスへ販売(当面は国内)

成果物の公開



ソースコード



モデル



開発ノウハウ

完全自動運転に向けたマルチモーダル基盤モデルの開発

実施者	Turing株式会社
概要	<p>完全自動運転を見据え、日本の運転環境に強い適応力を持つ、マルチモーダル(言語・画像・映像)基盤モデルを開発</p> <ul style="list-style-type: none"> • なお本事業はモデル開発が主眼で、本事業後に車体へ搭載し、改良を実施予定 <p>社会実装に向け、国内でのビジネス展開に加え、一部モデル・ノウハウ等の成果物も公開・提供</p> <ul style="list-style-type: none"> • 国内中心に、カメラ+AI方式の自動運転機能を有するEVを販売 • 開発で得られた成果物(ソースコード・モデル・開発ノウハウ)を公開

実施内容

- ① マルチモーダル学習基盤構築
 - 日本語データセットの構築・学習
 - 視覚モデルと結合
- ② 運転ドメインへの適合
 - 国内走行データに基づく画像・言語のデータセットを作成・追加学習
- ③ 分散環境による大規模学習・精度向上
 - 学習高速化、より大規模なパラメータサイズモデルへ対応

開発される基盤モデル

- 完全自動運転に向け、日本の運転環境に強い適応力を持つ、マルチモーダル基盤モデルを開発
- 日本語LLMをベースとし、画像・映像(1万時間分の走行動画など)と融合
 - なお本事業はモデル開発が主眼で、本事業後に車体へ搭載し、改良を実施予定



社会実装の方法

- 自社ビジネスとしての展開
- カメラ+AI方式の自動運転機能を有するEVの販売(当面は国内中心)

成果物の公開



ソースコード



モデル
一部のみ



開発ノウハウ

多様な日本語能力の向上を目指した公開型の基盤モデル開発

実施者	国立大学法人東京大学
概要	<p>民間企業・研究者・学生等の混成8チームが、小規模だが多様な基盤モデルを開発</p> <ul style="list-style-type: none"> 基盤モデル開発は技術黎明期であり、多様な開発手法を試行錯誤することで効率的な手法を探索 <p>開発を通じた開発者育成とともに、開発したモデル・開発ノウハウも公開し、日本の開発力底上げに寄与</p> <ul style="list-style-type: none"> 100名の開発者が本事業を通じて基盤モデル開発を実際に経験 開発で得られた成果物(ソースコード・モデル・開発ノウハウ)を公開

実施内容

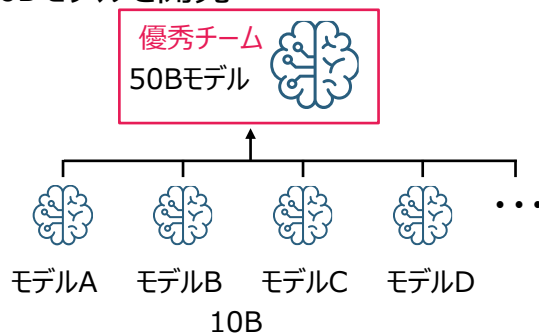
- ① 8チームそれぞれの10Bサイズの構築
- ② ①で最優秀チーム+若干名による50Bサイズの構築
- ③ 上記を通じた開発者の育成
- ④ 開発ノウハウの公開
- ⑤ 社会実装支援

開発される基盤モデル

民間企業・研究者・学生等の混成8チームが、小規模だが多様な基盤モデルを開発

- モデルの深さと幅の比率探索
- 最適なモデルの構造の探索
- LLMの精度向上につながるトークナイゼーションの改善 等

その後、8チームノウハウを集結させた50Bモデルを開発



社会実装の方法

成果物の公開



ソースコード



モデル
9つ全てのモデル



開発ノウハウ

関連スタートアップでの事業化

- 新規スタートアップの育成
- 参加者が所属するスタートアップでの活用
- 株式会社松尾研究所での個社・業界固有のLLM開発受注

100B/1Tパラメータからなる大規模マルチモーダル基盤モデルの構築

実施者	株式会社Preferred Elements
概要	<p>本開発事業において2つのモデルの開発、検証を行う。</p> <ul style="list-style-type: none"> 日本語性能に優れ、言語・画像・音声に対応したマルチモーダル100Bモデルの開発 1Tパラメータの言語モデルの事前学習の検証 <p>社会実装に向けたビジネス展開に加え、一部モデル・ノウハウ等の成果物も公開・提供</p> <ul style="list-style-type: none"> 本開発事業で得られた成果物(モデル・開発ノウハウ)の一部を公開 自社ビジネスとして、国内外でのAPI・ライセンスの提供、付帯するビジネスを実施

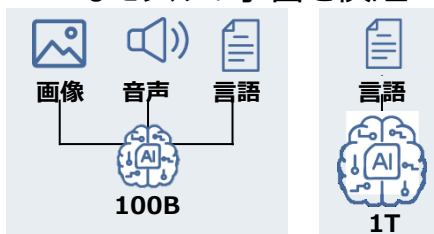
実施内容

- ① 100B モデルの事前学習・追加学習
- ② 指示学習
- ③ 画像モーダル向けの事前学習・追加学習
- ④ 音声モーダル向け追加学習
- ⑤ 1Tモデルの事前学習の検証



開発・検証する基盤モデル




- 次の2モデルを開発
- 100Bのマルチモーダルモデル
 - 言語・画像・音声に対応
 - 一部タスクで世界最高レベルの性能
 - 1Tの言語モデル
 - グローバルレベルでも大規模なモデルの学習を検証



社会実装の方法

- 自社ビジネスとしての展開
- 基盤モデルのAPI提供
 - 基盤モデルのライセンス提供
 - 基盤モデル及びAPIに付帯するビジネス（エンジニアリング・コンサルティング）

成果物の公開

 ソースコード (ファインチューニング用など)	 モデル 事前学習済 100Bモデル ウェイト	 開発ノウハウ マルチモーダル 化及び1Tモデル 学習
---	---	---