

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](#)

[モデル組み合わせによる日本語大規模基盤モデル開発と半自動データ作成フレームの構築](#)

[End-to-End音声基盤モデルの開発](#)

[論理推論を可能とする大規模言語モデルの研究開発](#)

[特化型モデル開発のためのモデルの小型化](#)

[動画生成AI基盤モデルと動画生成AIプラットフォームの開発](#)

[日本のアニメ産業活性化の為にアニメ分野特化型基盤モデル開発](#)

[生成AI 基盤による非定型帳票の革新と自律促進](#)

[薬学分野・製薬業務に特化したLLMの開発](#)

[リアルタイム音声基盤モデルの開発と日本市場における実用化](#)

[食品・流通小売領域の専門知識を取り込んだマルチモーダルな大規模モデル開発](#)

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

(g3) 競争力ある生成AI基盤モデルの開発

[世界最大規模の高品質データセットの構築およびそれを用いた大規模言語モデルの開発](#)

[AI創薬を実現させる分子情報特化基盤モデル開発](#)

[完全自動運転に向けた身体性を持つマルチモーダル基盤モデルの開発](#)

[都市時空間理解に向けたマルチモーダル基盤モデルの開発](#)

[地域気候サービスのための生成AI基盤モデルの開発](#)

[日本のカスタマーサポートのための高品質AIエージェントモデルの開発](#)

[ハルシネーションを抑止したドキュメント読解基盤モデルの構築](#)

[ユーザー意図を反映する選択的編集能力を備えたVision系基盤モデルの開発](#)

[創薬を加速する遺伝子発現量の基盤モデル開発](#)

[日本語とソフトウェア開発に特化した基盤モデルの構築](#)

[企業の知の結晶である様々なドキュメント群を読み取るマルチモーダルLLMの開発](#)

[観光用産業用向け405B LLM/基盤モデル開発](#)

(g6) A I 用計算資源の高度化に向けた研究開発

[多様なAI半導体の活用と計算資源の高効率化に関する研究開発](#)

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

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

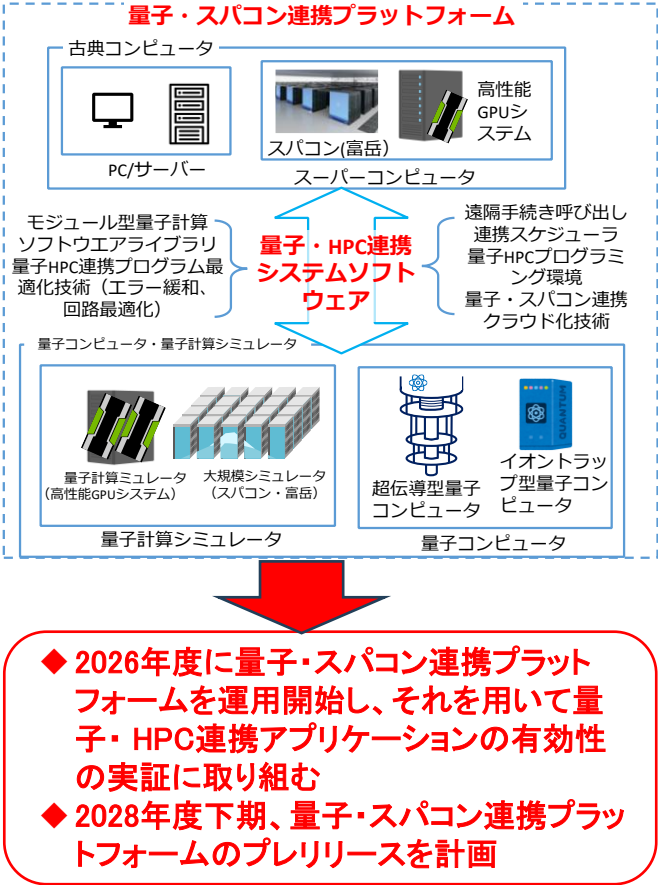
- 量子コンピュータは、従来のコンピュータと全く異なる原理で動作し、情報処理速度の劇的な高速化が期待されるが、現時点では、規模拡大と計算結果の誤り訂正の両立が困難であり、量子コンピュータ単独での実用化には時間を要する見込み。
- 一方で、デジタル化の進展により、情報処理能力の向上が急務であり、量子コンピュータの活用がいち早く求められているところ、古典コンピュータを組み合わせることで活用することが有望視されている。
- 本事業では、世界に先駆けて、量子コンピュータとスパコンを連携利用するためのソフトウェアやプラットフォーム、アプリケーションを開発・構築し、ポスト5G時代で提供されるサービスとして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



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】

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

プロセス微細化 (5nm以下予定)

低電圧化

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

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）

⇒ **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**

- 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

【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**

資源 ↑ 時間 →

最適化

効率化

効率化により、より多くのワークロードを実行可能に

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

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

開発される基盤モデル

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

LLMと外部データを結び、外部データに関連する文章を作成

膨大な外部データ

LLMが外部データと関連しながら必要なアクションを考案・実行

社会実装の方法

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	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

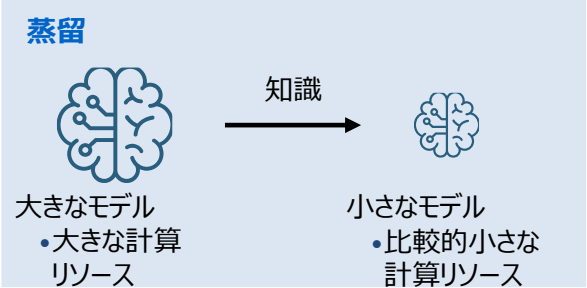
- ① 蒸留
- ② MoE(Mixture of experts)
- ③ 強化学習

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

蒸留



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	- 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

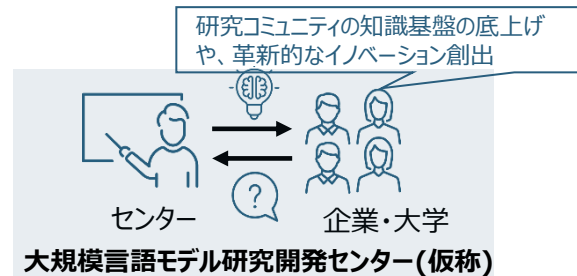


モデル



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	ストックマーク株式会社
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

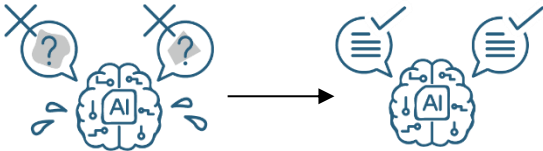
実施内容

- 事前学習・評価
- 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
- ②のデータを用いた事前学習・評価
- 時事話題等の追加事前学習・評価
- 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	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 - 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

- 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
- 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
- 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



画像/言語/映像を学習

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	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 - 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 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 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

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

優秀チーム
50Bモデル



...

モデルA モデルB モデルC モデルD

10B

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	本開発事業において2つのモデルの開発、検証を行う。 - 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 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 - 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

開発・検証する基盤モデル

社会実装の方法

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

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

画像

音声

言語

100B

言語

1T

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

モデル組み合わせによる日本語大規模基盤モデル開発と半自動データ作成フレームの構築

実施者	株式会社ELYZA
概要	日本語処理能力の高い高性能汎用基盤モデルを開発 • Mixture of Experts（MoE）、日本特有のデータの学習、フィードバック学習などを重点的に実施することで、推論効率の向上、日本特有の知識への回答精度の向上、ハルシネーションの抑制を目指す。 • 上記を通じ、開発した汎用基盤モデルの実用化および領域特化モデルの開発に繋げる。 社会実装に向け、自社プロダクトへの組み込み・API提供を進め、また一部モデル・ノウハウ等の成果物も公開・提供

実施内容

1 事前学習

2 指示学習

3 フィードバック学習

4 評価データセットの整備

5 コミュニティへの貢献

6 成果物の公開

開発される基盤モデル

グローバルモデルに匹敵する、日本語処理能力の高い汎用基盤モデルを開発

- MoE： 推論効率を向上
- 日本特有のデータの学習・フィードバック学習： 日本特有の知識への回答精度の向上・ハルシネーションの抑制

上記モデルの開発を踏まえ、領域特化モデルの開発に繋げる

汎用モデルの実用化
(本事業)

特化モデル開発
(本事業以降)

社会実装の方法

自社ビジネスとしての展開

- 自社既存プロダクトへの組み込み・API提供
- パートナー企業と共同で拡販

成果物の公開

ソースコード

評価などに関して極力公開

モデル

事前学習済み

開発ノウハウ

事前学習部分など

End-to-End音声基盤モデルの開発

実施者	株式会社Kotoba Technologies Japan
概要	End-to-Endで音声を入出力できる基盤モデルを構築、開発のインフラを整える - 大規模音声データセットを用いて、最大規模のEnd-to-End音声基盤モデルを学習 - 音声基盤モデルの評価パイプラインを整備 モデル学習のノウハウ / 学習済みモデルの一部を公開する、社会実装に向けた国内でのビジネス展開を行う - モデル学習のノウハウをテックブログ、コードベースなどの形で公開 - 学習済み小規模モデルも安全性を考慮した上で公開 - 開発した音声基盤モデルのAPIとしての公開、商用パッケージング

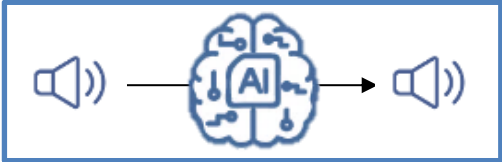
実施内容

- データセットの構築
- 音声基盤モデル開発
- 学習モデル評価
- 知見の公開
- コミュニティへの貢献
- API化、商用パッケージング

開発される基盤モデル

音声AIの汎用性の強化に向けて、End-to-Endの音声基盤モデルの開発をスケールアップ：

- ・ **学習データ**：日英の大規模音声データセットを数十万時間活用
- ・ **モデル**：最大70億パラメータの音声モデルとしては最大規模のモデルを開発



音声の入出力をEnd-to-End基盤モデルで行う。

社会実装の方法

自社ビジネスとしての展開

- ・ **API提供**：開発した音声基盤モデルを企業・デベロッパーに向けて商用提供
- ・ **商用パッケージング**：音声基盤モデルをビジネスに向けてパッケージングして販売

成果物の公開



ソースコード
追加学習・推論用のコードなど



モデル
小規模モデルの重みを学術機関に向けて公開、企業にも条件付きで公開



開発ノウハウ
テックブログなどを通じてノウハウを公開(前処理、評価、分散並列学習など)

論理推論を可能とする大規模言語モデルの研究開発

実施者	富士通株式会社
概要	法務・金融・医療など規制規則への準拠が必要な業界でも生成AIを利用可能にするため、知識処理技術ナレッジグラフの生成と推論に特化した、以下2つの大規模言語モデル（LLM）を開発 ・ナレッジグラフ生成LLM）自然言語テキストからナレッジグラフを生成する大規模言語モデル ・ナレッジグラフ推論LLM）ナレッジグラフに基づいて質問回答を論理推論する大規模言語モデル 社会実装に向け、自社ビジネスに関連させた展開を図り、モデル・開発ノウハウ等の成果物も公開・提供 ・本開発事業で得られた成果物の一部を、Hugging Face/GitHub等にてOSS公開

実施内容

1 データセットの構築

事前学習用に、自然言語テキスト⇔ナレッジグラフの対訳コーパスを整備
ナレッジグラフの生成タスクと推論タスクの指示学習データを作成

2 アーキ・ハイパラの選定

3 LLMの開発と精度向上

上記学習データを用いて7B～46.7B級(MoE構成含む)のLLMからベースとなる1つを選定し、事前学習・指示学習を実施

4 成果物の公開

開発される基盤モデル

・ナレッジグラフ生成LLM

関係抽出ベンチマーク（英語：ReDocRED, 日本語：JacRED）で世界最高性能+10ポイントを達成

自然言語文書

→



→

ナレッジグラフ

・ナレッジグラフ推論LLM

マルチホップQAベンチマーク（英語：HotpotQA, 日本語：JEMHopQA）と法的判断予測ベンチマークで世界最高性能+10ポイントを達成

質問文

→



→

ナレッジグラフ

→

回答文+論理推論

社会実装の方法

自社ビジネスとして展開

- ・**Fujitsu Kozuchi**) 自社AIプラットフォームへの搭載により、企業・一般向けに広く試用環境を提供
- ・**Fujitsu Uvance**) 各種ビジネスオフリングに組み込み、グローバルにソリューション展開

成果物の公開



ソースコード
モデル利用のためのコード



モデル



開発ノウハウ

特化型モデル開発のためのモデルの小型化

実施者 株式会社ABEJA

- 概要**
- 本開発事業において、以下要件を満たし実運用を見据えた大規模言語モデルの開発を行う
 - 特定タスクにおいて高性能で実用可能
 - モデルのパラメータサイズの規模を抑える（50B以下及び10B以下の2つのモデル）
 - 開発プロセス及び一部学習用ソースコード・モデルの成果物も公開・提供し、コミュニティの活性化・発展にも貢献する

実施内容

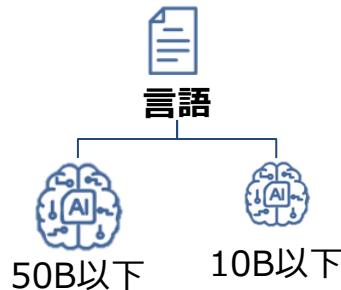
- 1 モデルの継続事前学習（50B以下/10B以下両方）
- 2 特定タスクを見据えた指示学習用データセットの構築
- 3 特定タスク重視の指示学習
- 4 特定領域※向けの指示学習



開発される基盤モデル

次の2つのモデルを開発

- 50B以下の言語モデル
 - 特定タスク※において、世界トップの性能
- 10B以下の言語モデル
 - 実運用可能な性能と運用コストの両立



※特定タスク / 特定領域:
ビジネス支援におけるタスク・ドメイン領域を想定

社会実装の方法

- 自社ビジネスとしての展開
 - ABEJA Platformへの搭載
 - ABEJA Platformを通じて各顧客企業へ導入
 - ABEJA Platformを基盤とした各顧客企業業務のLLM連携による生産性向上

成果物の公開


ソースコード


モデル
※事後学習
モデルは非公開


開発ノウハウ

動画生成AI基盤モデルと動画生成AIプラットフォームの開発

実施者 株式会社AideaLab

- 概要**
- 本開発事業において3つの動画生成AI基盤モデルを開発
 - 日本語汎用軽量動画生成AI基盤モデルの開発
 - 日本語汎用動画生成AI基盤モデルの開発
 - 日本語アニメ生成AI基盤モデルの開発
 - モデル公開や動画生成AIプラットフォームを通じて、自社ビジネスを展開することで社会実装
 - 日本語汎用軽量動画生成AI基盤モデルやソースコードの公開と開発ノウハウの共有
 - 日本語アニメ生成AI基盤モデルによるコンテンツ制作支援サービスの提供と開発ノウハウの共有

実施内容

- 1 軽量で汎用の動画生成AIの開発
- 2 汎用の動画生成AIの開発
- 3 アニメ用動画生成モデル開発

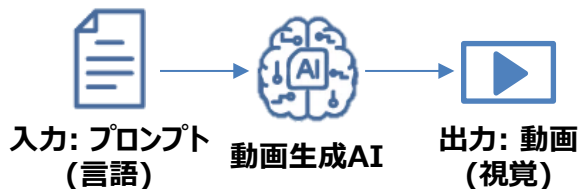
開発される基盤モデル

- 軽量で汎用の動画生成AI
 - 一般的なGPUでも扱える軽量の動画生成AIの公開
- 汎用の動画生成AI
 - 一般の動画生成AI
- アニメ用の動画生成AI
 - 世界最高レベルの映像を実現

社会実装の方法

- 自社ビジネスとして展開
 - 動画生成AIプラットフォームによるユーザーへの提供
 - 日本語汎用動画生成AI基盤モデルのAPI提供
 - 日本語アニメ生成AI基盤モデルのAPI提供
 - 基盤モデルによるソリューション提供

成果物の公開




ソースコード
(追加学習用 等)


軽量モデル


開発ノウハウ
(動画生成AIの
開発時の注意点)

日本のアニメ産業活性化の為にアニメ分野特化型基盤モデル開発

実施者 AiHUB株式会社

- 概要**
- 開発内容
 - 次世代のAIモデルである生成AIオーケストレーションと親和性の高い且つアニメ分野特化の基盤モデルの開発
 - 社会実装（事業化）
 - 基盤モデルを利用したサービスをSaaS形式でアニメ制作の中心である日本のアニメ制作会社向けにリリース

実施内容

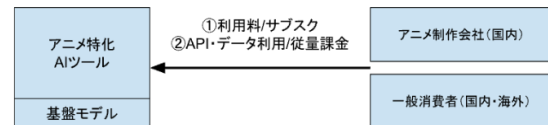
- 1 原理実証モデルの構築
- 2 基礎概念モデルの構築
- 3 追加学習（アニメ特化）

開発される基盤モデル

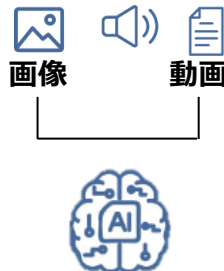
- アニメ分野に特化した基盤モデル
- 権利処理されたアニメ画像生成で実用レベルでの性能
 - キャラ設計（イメージパターンの創出）
 - 3Dモデルからアニメ調にコンバート
 - 中割りの生成
 - 背景の生成
 - 線画からの自動着彩

社会実装の方法

- 自社ビジネスとしての展開
 - 基盤モデルを利用したアニメ事業者に向けアニメ特化生成AIサービスの提供



成果物の公開



- | | | |
|---|---|---|
|  |  |  |
| 一部ソースコード
（限定公開） | 一部モデル
（限定公開） | 開発ノウハウ
（全体公開） |
| ※国立情報学
研究所 等 | ※国立情報学
研究所 等 | |

生成AI 基盤による非定型帳票の革新と自律促進

実施者 AI inside株式会社

- 概要**
- 本プロジェクトは、非定型帳票の処理効率を向上させるため、生成AIモデルにおけるLLM-SLMの2段階構成を採用。SLMが簡易な処理を担い、コスト削減と処理負担軽減を図る。また、強化学習による精度向上、SLMの自律的な蒸留技術、スケーラブルなインフラ構築、分散計算機能の研究開発を進め、限られたリソースでの最適化を目指す
 - また、社会実装に向け、サービスを多様な形態（Public Cloud/Private Cloud/On-premises）を提供し、開発ノウハウや評価セット等の成果物も公開する

実施内容

- 1 学習・評価データセット構築
- 2 自社LLMの精度改善
- 3 ベンチマークでの評価実施
- 4 スケーラブルなインフラ構築
- 5 社会実装に向けた分散計算機能の研究開発
- 6 評価データセット、ノウハウの公開

開発される基盤モデル

- 二段階生成AI (LLM-SLM) 推論アーキテクチャ及びSLM自律的に蒸留技術の開発を行う
- 運用しながら低コストで高品質なSLMを作る仕組みを確定する
- 生成AIの運用基盤を確保しやすく、分散計算（学習・推論）にしやすい構成を開発



社会実装の方法

- 自社サービスとの連携を図り、ノウハウについては特許を申請し、公開・提供
 - 自社サービスに研究開発したモデルを組み込むことによる社会提供
 - インフラチューニングを含めて最適化済みのサービスを多様な形態（Public Cloud/Private Cloud/On-premises）を提供予定

成果物の公開



ソースコード
モデル利用ための
コードのみ



評価データ
セット



開発ノウハウ

薬学分野・製薬業務に特化したLLMの開発

実施者	株式会社EQUES
概要	【開発内容】 ・薬学分野のデータセット構築および評価ベンチマークの作成 ・薬学分野や製薬業務におけるタスク遂行能力に優れた日本語LLMの開発 【社会実装（事業化・実用化・成果物公開）】 ・本開発事業で得られた成果物（ベンチマーク・モデル・開発ノウハウ・論文執筆）の積極的な公開 ・自社ビジネスとして国内の製薬会社を中心にアプリケーション・AIソリューションの提供

実施内容

- 1 データセットの構築・タスク設計
- 2 手法検証
- 3 モデル本開発
- 4 モデル性能評価実験
- 5 アプリケーション性能検証

開発される基盤モデル

- 薬学・製薬に焦点を当てた
軽量なドメイン特化LLM
- ・ 7～9Bの言語モデル
 - ・ 薬学関連ベンチマークで同規模の既存LLMを優越する性能
 - ・ 製薬業界における実際の業務を模した評価タスクにてGPT-4などに比肩する性能

社会実装の方法

- ・ 基盤モデルを活用したアプリケーションやAIソリューションの各社への提供
- ・ 基盤モデルを用いた共同研究（例：セキュア環境のみで実施可能なもの）

【成果物の公開】



リアルタイム音声基盤モデルの開発と日本市場における実用化

実施者 株式会社Kotoba Technologies Japan

- 概要**
- ・ 当事業では、音声基盤モデル、すなわち音声データをEnd-to-Endにリアルタイムかつ流暢にやりとりできる基盤AIモデルの開発を進める。大規模な日本語・多言語音声データを活用して、8Bの汎用的な能力を待つ音声基盤モデルを学習する
 - ・ 学習した音声基盤モデルは、音声Chatbotや音声翻訳ツールのようなアプリケーションの形でplaygroundやエンタープライズの顧客候補に向けたプレビューを行い、商業化への道筋を示す

実施内容

- 1 データセット構築 / TTS (Text-to-Speech) モデル学習
- 2 初期実験
- 3 音声基盤モデル学習
- 4 アプリケーションへの社会実装
- 5 日本AIコミュニティへの貢献

データ構築と開発される基盤モデル

- 日本語を中心とした音声データを、合成データも活用しながら60万時間分の高品質音声データセットを構築
- ・ 日本語TTSモデルを学習・活用して20万時間合成、その他のリソースから日本語データを20万時間を収集、クリーニング
 - ・ 英語を中心とした20万時間分の高音質データを収集、クリーニング

上記のデータに8Bの音声基盤モデルを学習させて、音声Chatbot等の汎用的対応に加えて、高速推論・Two-Stream推論等を可能にする



音声の入出力をEnd-to-End基盤モデルで行う

社会実装の方法

- ・ 自社ビジネスとしての展開
 - TTSのPlaygroundや8B音声基盤モデルを音声Chatbotや音声翻訳ツール等の形でプレビュー公開
 - エンタープライズの顧客候補に対しては音声基盤モデルをユースケースに特化させてプレビュー提供

成果物の公開



前処理モデル

前処理パイプラインに使うASRモデルやラベリングモデル等はHuggingFaceで公開



開発ノウハウ

音声基盤モデルの訓練や推論に関する知見を適宜テックブログやホワイトペーパーを通して公開

食品・流通小売領域の専門知識を取り込んだマルチモーダルな大規模モデル開発

実施者 NABLAS株式会社

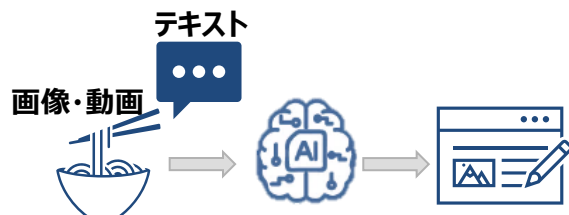
- 概要**
- 日本特有かつ流行をおさえた食品に強い生成 AI 技術の開発
 - 日本の食品業界特有の知識を高度に扱える大規模視覚言語モデル (Large Vision-Language Model: LVLM) の開発
 - 画像、動画、テキストという異種データを同時に処理できるグローバルなベースラインモデルに対抗しうるモデルを構築
 - 社会実装に向け、サービス化するとともに、一部モデル・ノウハウ等の成果を公開
 - 食品業界への事業展開
 - 技術報告書、一部のデータセット・モデル、技術文書、ドメイン特化の指示学習データセットの作成方法の公開

実施内容

- 事前学習①
 - LLaVA-Pretrain上での事前学習
- 事前学習②
 - モデルの改善
- 指示学習①
 - 汎用的なモデル構築
- 指示学習②
 - 独自データセット上での学習
- 混合専門家モデルへの拡張
- 成果物の公開

開発される基盤モデル

- 日本の食品業界特有の知識を高度に扱える大規模視覚言語モデル (Large Vision-Language Model: LVLM)
- 事前学習を行い、パラメーター数の異なるLVLM数種を開発
 - 上記のMoE拡張版を開発
 - 指示学習データセットの異なるLVLMを開発
 - 汎用型
 - 領域特化型



社会実装の方法

- サービス化
 - アプリケーション開発
 - 食品領域企業向けに展開

成果物の公開

- 技術報告書
- 事前学習データセット
- 一部の指示学習データセット
- 事前学習済みのモデル
- 一部の指示学習済みのモデル
- 得られた知見、ノウハウ


ソースコード


モデル


開発ノウハウ

世界最大規模の高品質データセットの構築およびそれを用いた大規模言語モデルの開発

実施者	株式会社Preferred Networks、株式会社Preferred Elements
概要	本開発事業においてデータ生成およびモデルの開発を行う - LLMを活用した世界最大規模の高品質な学習データを構築 - 構築したデータを含めフルスクラッチの事前学習および事後学習を行い、最大30Bパラメータ Mixture of Expertsモデルを開発 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 - モデルの重み、データ生成方法、モデル性能、学習時のノウハウを公開予定 - 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

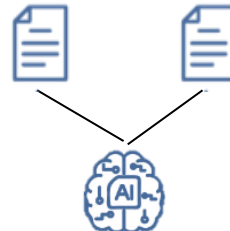
- ① 高品質なデータ構築
- ② 基盤モデル開発
- ③ ノウハウの共有および公開
- ④ 事業化



開発される基盤モデル

- MoEからなる30Bモデル
 - JMMLU、Jaster、JMT-BenchにおいてGENIAC1.0で構築したPLaMo-100Bを超える性能
 - JHumanEvalでllama 3 Swallow 70Bを超える性能

既存のデータセット LLMを活用した高品質なデータセット



30
B

社会実装の方法

- ノウハウの共有・公開
 - 学術研究用途などが行えるようなライセンス形態でのモデルの重みを公開予定
 - データの生成方法、学習時における工夫、学習結果、モデル性能などはブログ等で公開予定
- 自社ビジネスとしての展開
 - PLaMoシリーズのクラウドを介した提供
 - PLaMoシリーズのライセンス提供
 - PLaMoシリーズのファインチューニングなど改良開発



ソースコード



モデル



開発ノウハウ

AI創薬を実現させる分子情報特化基盤モデル開発

実施者 SyntheticGestalt株式会社

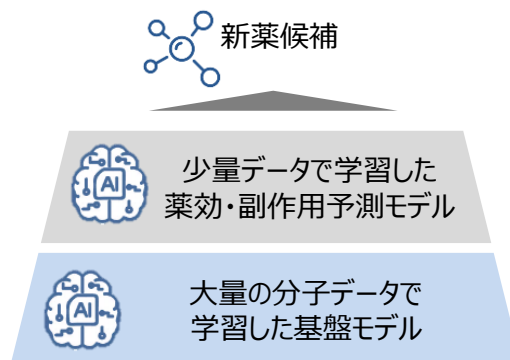
- 概要**
- AI創薬に特化した基盤モデル開発・検証を行う
 - 様々な創薬課題を解決する機械学習モデルを開発するための基盤モデルを開発する
 - 最大100億件の化合物の立体構造情報を学習データとして使用する
 - 上記基盤モデルを利用して薬効・副作用（Affinity及びADME-Tox）予測能の検証を行う
 - モデル検証のために実験を行い評価データを作成し、評価データは公開する
 - 本事業により開発される小型モデルも公開する

実施内容

- 分子の立体情報に特化した基盤モデルの開発
- 基盤モデルを利用して、公開ベンチマーク評価用に23種類と実利用データ評価用に27種類の応用モデルの開発及び評価
- Enamine社に外注してAI創薬用途の基盤モデル評価用データの作成
- 小型版基盤モデルの開発・公開
- 大手クラウドベンダーと提携して世界展開予定
- GENIACが主催するコミュニティへの参画

開発される基盤モデル

創薬用途のAI開発に貢献することを目指し、最大100億件の分子の立体構造を学習した基盤モデルを開発する。この基盤モデルによる低分子の薬効・副作用予測性能の向上を評価し、AI創薬課題の標準ベンチマーク全てに対して上位3位以内のランク入りを目指す



社会実装の方法

ビジネス展開

Google Cloud及びAWSのMarketplaceを通じてモデルを展開し、同社らと連携しながら世界の創薬・材料・化学企業に展開

成果物公開

作成した評価用データの全てと、分子基盤モデルの小型版を公開



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身体性を持つ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**
- 完全自動運転を見据え、「身体性」を持つマルチモーダル基盤モデルを開発
 - テキストや映像、センサーデータを統合的に理解し、運転環境における周囲の移動体や環境の変化をリアルタイムで予測するモデルを開発する
 - 社会実装については、完全自動運転に向けた開発を推進するが、そのための要素技術の一部については他領域でも利用可能な共通基盤として公開や提供を予定

実施内容

- 大規模マルチモーダルモデルの高度化
 - 視覚-言語データセットの構築・検証
 - 視覚-言語マルチモーダルモデルの学習
- 三次元情報を含む自律移動データセットの構築
 - 複数センサを用いた走行データの取得
 - 三次元情報アノテーションと言語表現の付与
- 身体性ある自動運転モデルの開発
 - マルチモーダルモデルの身体性拡張
 - 自動運転システムへの結合

開発される基盤モデル

- 300億パラメータ超の視覚/言語/行動**に対応した統合的なマルチモーダル基盤モデルを開発
- 視覚-言語の学習を軸としつつ、現実世界の理解および実際の運転タスクに緊密に連携
 - 自動運転タスクにおいて、各種ベンチマークで評価し、国際的に報告されている最高水準の性能を目指す

社会実装の方法

自社ビジネスとしての展開

- カメラ+AI 方式の自動運転システムを開発し、自動車メーカーやティア1メーカーへADASシステム（先進運転支援システム）として販売する



成果物の公開



都市時空間理解に向けたマルチモーダル基盤モデルの開発

実施者	ウーブン・バイ・トヨタ株式会社
概要	- City-LLM: 都市時空間理解に向けたマルチモーダル(言語・映像・センサ)基盤モデルの開発 - 実世界のドメインに対して、動画像内インスタンスレベルでの時空間情報への理解に注力する - 映像＋画像＋文字を有するデータセット6億を用い、7Bレベルのモデルを開発する - 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を公開 - Woven Cityでのリアルな都市環境での機能検証とビジネス検証の実施 - 国際ワークショップ, AI City Challenge (CVPR) にて比較検証を行う - 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)の一部公開

実施内容

- 1 高品質な大規模マルチモーダルデータセットの整備
- 2 分散学習基盤環境の構築
- 3 動画像と言語のコネクタ部分は、映像・画像・言語を用いた事前と事後によるフルスクラッチ学習を行う
- 4 LLMの言語モデルもファインチューニングを実施
- 5 Video LLMの国際ベンチマークデータセット (MVBench) で評価

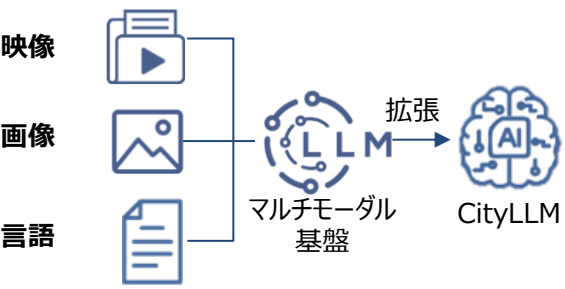
開発・検証する基盤モデル

時空間情報への理解を中心となるマルチモーダル基盤モデルの開発、特に、以下の2点に対して取り組む

- (1) インスタンスレベルの時空間情報を言語空間に整合する学習パイプラインの設計
- (2) マルチモーダルの入力に対して、出力を安定させるフィードバック機構

社会実装の方法

- Woven Cityでのリアルな都市環境での実装, 国際ワークショップ開催によるオープンイノベーションの促進, 自社ビジネスとしての展開
 - 交通安全分野でのサービス
 - 人の移動を支援するサービス
 - 実世界における検索サービス



成果物の公開

ソースコード
(モデル利用
ためのコードのみ)

モデル
(一部)

開発ノウハウ
(論文等)

地域気候サービスのための生成AI基盤モデルの開発

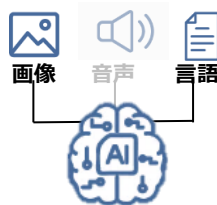
実施者	国立研究開発法人海洋研究開発機構
概要	本事業では、日本国内における特定地域や企業が効果的な気候変動対策を立案するための地域気候サービス特化型モデルの開発を行う。 特に日本語対応の言語モデルまたは視覚言語モデルをベースとし、気象・気候学の基礎から将来の温暖化に伴う気候変動予測情報の知識を獲得することによって、地域や企業において想定されるリスク情報や効果的な対策（適応策、緩和策）立案を行う。

実施内容

- 1 学習データ収集
- 2 ベンチマーク構築
- 3 モデルのチューニング
- 4 マルチモーダルRAG構築
- 5 評価・検証
- 6 公開
- 7 実用化／事業化
- 8 コミュニティの参画

開発される基盤モデル

- 以下の2モデル（8B程度）を開発
- ・ TCFDレポート生成用モデル
 - ・ リスク情報（移行リスク、物理的リスク）
 - ・ リスク対策（ガバナンス、戦略、リスク管理、指標・目標）
 - ・ 自治体の災害対策立案用モデル
 - ・ リスク情報（ハザード、脆弱性、暴露）
 - ・ リスク対策（防災行動計画等）



社会実装の方法

- ビジネス展開
- ・ モデル企業、モデル自治体との実証
 - ・ 基盤モデルのライセンス提供（TCFDレポート生成用モデルは有償、災害対策立案用モデルは将来的な公開も想定）
 - ・ ベンチマーク公開



ソースコード



モデル



開発ノウハウ

日本のカスタマーサポートのための高品質AIエージェントモデルの開発

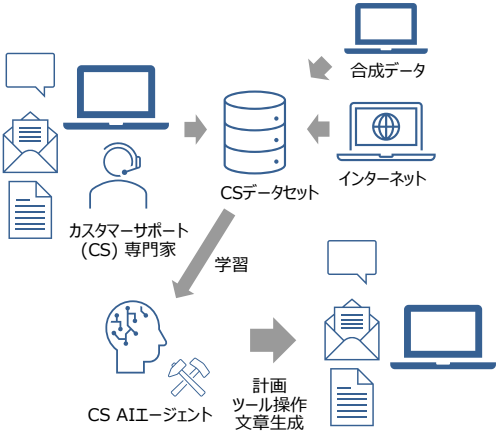
実施者	カラクリ株式会社
概要	- 本開発事業では主に以下の2つを実施する - 日本のカスタマーサポートに特化したベンチマークを構築する - 日本のカスタマーサポートに特化したエージェントモデル（70B相当）を構築する - 社会実装に向けたビジネス展開に加え、一部モデル・ノウハウ等も公開・提供 - 本開発事業で得られたモデル、開発ノウハウを公開 - 自社ビジネスとして、自社サービス内への組み込みを実施

実施内容

- 1 カスタマーサポート特化のベンチマークの作成
- 2 カスタマーサポート特化のモデルの作成
- 3 モデル・ベンチマーク・ノウハウを公開

開発される基盤モデル

- カスタマーサポート特化AIエージェントモデル
- カスタマーサポートのための知識を学習
 - 課題に対して適切な計画を立て、ツール操作をこなすことができるモデル
 - 今回作成するカスタマーサポートベンチマークでは世界最高の性能



社会実装の方法

- 基盤モデルの公開
 - 商用利用可
- 開発ノウハウの公開
 - 技術ブログ
 - イベント登壇
- ソースコードの公開
 - 利用しやすいライブラリを公開
- 既存の当社サービスへの組み込み
 - 年間延べ1000万人が利用する当社サービスを通して提供

公開する情報


ソースコード


モデル


開発ノウハウ

ハルシネーションを抑止したドキュメント読解基盤モデルの構築

実施者 ストックマーク株式会社

- 概要**
- ・ 実ビジネスで活用される、文字だけでなく図表や概念図を含む設計書・提案書等の創造性の高い複雑なドキュメントを読解できるマルチモーダル（言語・視覚レイアウト）基盤モデル（100Bサイズ）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
 - PaaS/API化し他社へ販売、情報収集・検索SaaSの既存自社サービスへの組み込み・販売
 - 開発した基盤モデルやマルチモーダル基盤モデルの公開、開発ノウハウの提供

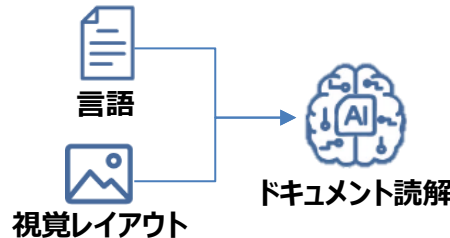
実施内容

- 100Bサイズ基盤モデルの事前学習の事前検証
 - ・ 良質なハイパーパラメータの特定 等
- 100Bサイズ基盤モデルの事前学習（2Tトークン程度）
 - ・ ビジネス領域に力点が置かれた学習データセットで学習
- マルチモーダル学習によるドキュメント読解基盤モデル構築

開発される基盤モデル

文字や図表・概念図を含む設計書・提案書等の複雑なドキュメントを読解できるマルチモーダル基盤モデル

- ・ 2Tトークン程度の事前学習
- ・ 独自作成する50K程度のドキュメント読解用教師データ等を用いたマルチモーダル学習
- ・ 複雑な図表を含む独自評価セットでGPT-4o以上の読解性能 + ChartQA等の一般ベンチマークでGPT-4oと同等程度のスコア



社会実装の方法

- ・ 自社ビジネスでの展開
 - PaaS/API化によって他社サービスへ販売
 - 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

成果物の公開


ソースコード


モデル
事前学習済
100Bモデル
ウェイト


開発ノウハウ
100B事前
学習・マルチ
モーダル学習

ユーザー意図を反映する選択的編集能力を備えたVision系基盤モデルの開発

実施者	株式会社データグリッド
概要	- 選択的編集性能の高いVision系基盤モデルの開発 - 実ビジネスに応用する上で不可欠な選択的編集能力が高い汎用動画・画像生成基盤モデルと、生成コンテンツを検知するディープフェイク検知モデルを同時構築 - 社会実装に向け、自社プロダクトへの組み込み及びAPIプラットフォームの構築を進め、モデル・開発ノウハウ等の成果物も公開・提供 - 本開発事業で得られた成果物の一部をHugging Face、GitHub、テックブログ等にて公開

実施内容

- 1 データセット整備
- 2 モデル構築
- 3 性能評価
- 4 コミュニティへの貢献
- 5 成果物の公開

開発される基盤モデル

- A. 動画画像生成基盤モデルの開発
- 汎用動画生成基盤モデルの開発
 - オープンモデルの中で、FVD等の指標でSOTAを達成できる基盤モデルの開発
 - 汎用画像生成基盤モデルの開発及びドメイン特化基盤モデルの開発
 - 選択的編集能力（CLIP, DISTs 等）でSOTAを達成する基盤モデルの構築
 - 独自のデータセットで製造業特化モデルを構築
- B. ディープフェイク検知基盤モデルの開発
- 基盤モデルから得られる特徴量を生かした独自モデルを構築



生成



検知



社会実装の方法

- 自社ビジネスとしての展開
 - 自社プロダクトである製造業向けAIデータ生成基盤「Anomaly Generator」への組み込み
 - APIプラットフォームの構築及びパートナー企業との拡販

成果物の公開



ソースコード

推論コードのみ



モデル

一部モデルを公開

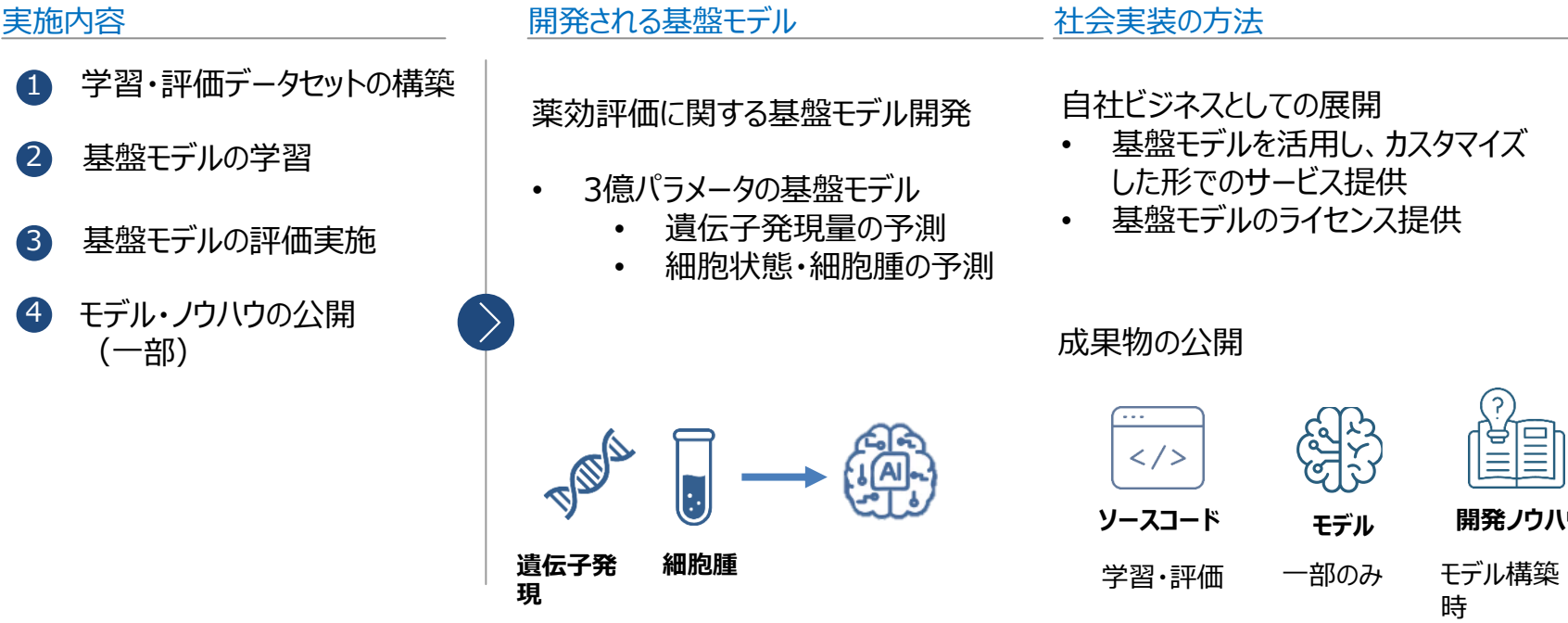


開発ノウハウ

モデル構築時

創薬を加速する遺伝子発現量の基盤モデル開発

実施者	株式会社ヒューマノーム研究所
概要	創薬の成功率を向上に向けた、分野特化型基盤モデルの開発 - 遺伝子発現量の計測実験の結果を活用した、3億パラメータ相当の基盤モデルの構築 社会実装に向け、国内でのビジネス展開に加え、一部データ・モデル・ノウハウ等の成果物も公開・提供 - 本事業で開発した基盤モデルをカスタマイズした形でサービス提供 - 研究開発で得られた成果物（ソースコード、モデルの一部、開発ノウハウ）を公開



日本語とソフトウェア開発に特化した基盤モデルの構築

実施者 フューチャー株式会社

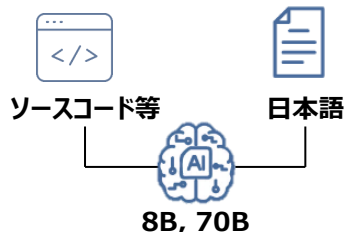
- 概要**
- 継続事前学習により日本語とソフトウェア開発に特化した基盤モデルを構築
 - 世界最高峰の日本語×ソフトウェア開発能力を目指す
 - コードレビュー・詳細設計・単体試験・ドキュメント生成等、ソフトウェア開発におけるコーディング以外の周辺タスクについても性能向上を目指す
 - 情報教育への応用も見込み、IT人材育成へも貢献
 - 社会実装に向け、ビジネス展開に加え、一部モデル、新規構築予定の評価データ、開発ノウハウを公開

実施内容

- 日本語とソフトウェア開発データを用いた継続事前学習
- ソフトウェア開発に関連した日本語Instruction-tuningデータの構築
- 上記データを用いたInstruction-tuning
- ソフトウェア開発に関連した日本語評価データの構築
- 一部モデル及び評価データ、ノウハウの公開
- 社会実装に向けた事前検討

開発される基盤モデル

- 日本語とソフトウェア開発に特化した8B, 70Bの基盤モデルを構築
- 従来のモデルを大きく上回る日本語×ソフトウェア開発能力
 - これまでの基盤モデルが得意としていたコード補完に加えて、コードレビュー・詳細設計・単体試験・ドキュメント生成へも対応
 - 情報教育への応用に向けて、教育的なコードレビュー生成にも取り組む
 - 日本語を用いたソフトウェア開発能力評価データについても新規構築



社会実装の方法

- 基盤モデルを活用したソフトウェア開発支援システムの構築
- 基盤モデルのライセンス提供
- 基盤モデル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）
- 会津大学と連携し、Aizu Online Judgeへの教育支援システム実装

成果物の公開

		
ソースコード	モデル	開発ノウハウ
(評価用ソースコード)	(8Bモデル)	(新規構築評価データ含む)

企業の知の結晶である様々なドキュメント群を読み取るマルチモーダルLLMの開発

実施者	株式会社リコー
概要	- 日本のモノづくりの知の継承の危機を解決する事をめざし「企業の知の結晶である様々なドキュメント群を読み取るマルチモーダルLLM」を開発 - マニュアル等の様々な図表を読解可能なマルチモーダル性能 - 製造現場で使われる長文のマニュアルや指示書を理解するためのロングコンテキスト処理能力 - 企業特有のドキュメント群に容易に適用可能なカスタマイズ性能 - 社会実装とともに、モデル・開発ノウハウ・ベンチマークを公開することで今後の基盤モデル開発に貢献

実施内容

- 1 マルチモーダル学習基盤構築
- 2 LMMアーキテクチャ検討
- 3 LLMを用いた学習データ生成
- 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

- ドキュメント群の読み取りに適したマルチモーダルモデル
- 高解像度対応
 - ロングコンテキスト対応
 - 高いカスタマイズ性



画像



言語



プライベートLMM対応
/カスタム対応

社会実装の方法

- 自社ビジネスとして展開
 - マルチモーダルモデルのAPI提供
 - マルチモーダルモデルのライセンス提供
 - マルチモーダルモデルに付帯するビジネス（カスタマイズ・コンサルティング・展開）

成果物の公開



ソースコード
一部のみ



モデル
顧客データを
含むモデルは除く



開発ノウハウ

観光用産業用向け405B LLM/基盤モデル開発

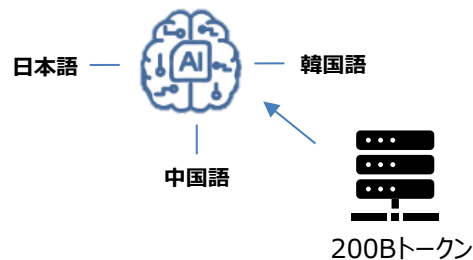
実施者	株式会社ユビタス、株式会社Deepreneur
概要	東アジア言語（日中韓）に強い多言語LLM基盤モデルを開発公開 • Llama3.1 405Bをベースに日中韓3言語の事前追加学習（計200Bトークン） & 命令チューニング • 訪日観光客・外国人労働者向けLLMアプリケーション開発を促進 基盤モデルの社会実装事例として観光産業向けLLMを開発・収益化 • 観光スポット情報等を活用したファインチューニング & RAGデータ連携 • 自治体・カーナビ等向けに多言語案内アプリケーションを提供

実施内容

- 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 Llama3.1 8B、70B、405Bとパラメータを増やしプレトレーニングを実施
- ⑤ ベンチマークでの評価実施
- ⑥ 観光データを一部ファインチューニング
- ⑦ 観光向けアプリケーションに搭載し、国内観光業の発展に貢献

開発される基盤モデル

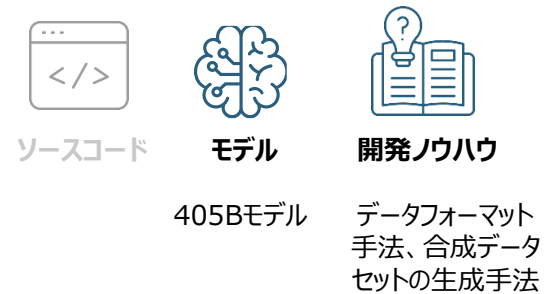
日本語・中国語・韓国語・英語による
情報アクセスに強い基盤モデルを開発



基盤モデルをベースに観光産業向け
案内アプリケーションLLMを実装

社会実装の方法

成果物としてモデルとデータセット
を公開



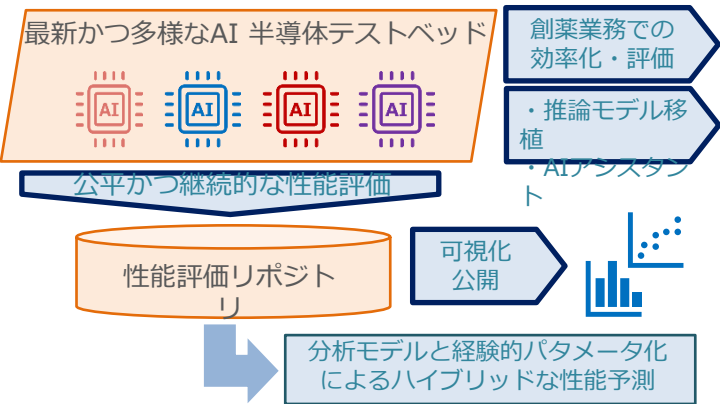
自治体向けWeb案内アプリケーション、
交通手段のアプリケーションを
提供・事業化

多様なAI半導体の活用と計算資源の高効率化に関する研究開発

実施者	国立研究開発法人産業技術総合研究所、1FINITY株式会社、株式会社AI福島、株式会社ELEMENTS、富士通株式会社、株式会社デブコシステムズ、株式会社RUTILEA、株式会社ゼウレカ
概要	AIの開発力・供給力を支える国内計算基盤の高度化のため、 <u>多様なAI半導体の評価と計算資源の高効率化を実現する技術</u> を研究開発する。高性能化や低消費電力が期待されている複数の新興AI半導体からなるテストベッドを整備し、 <u>その性能や省電力性、AI開発の利便性、運用性等の多面的な評価</u> を行い、用途に応じた利活用指針を明らかにする。また、それらAI半導体を含む計算資源の高効率化として、 <u>AIワークロードの実行性能予測や高効率・高性能推論システム</u> を開発する。

【①多様なAI半導体の評価と効率的な利用技術の研究開発】

- 多様かつ最新のAI半導体の特性・優位性を明らかにする
 - テストベッド構築と公平なAIベンチマークセットの開発
 - 実業務（創薬ワークロード）での効率的な利用技術の開発
 - 多様なAI半導体での推論モデル移植・実行技術及び生成AIによるアシスタント機能の高度化
- ### ②-1) 実行性能予測に基づく資源利用の効率化
- 学習／推論の高精度な実行性能予測技術の開発
 - 学習／推論処理の継続的な性能評価と性能予測用データセットの高度化



②-2) 高効率・高性能 推論システムの研究開発

- 入力データ選択制御：多種多様なデータソースからの入力データ量の削減とクエリ処理の効率化・高速化技術の開発
- LLM推論API：日本語とデータ操作言語を組合せた入力に対し投機的デコーディングによる適切な推論エンジン選択技術の開発
- 構成管理ソフト：需要変動に基づく、推論基盤モデル向け計算資源構成の動的適正化技術の開発
- スケジューラ：AI半導体の利用効率を上げる高度なスケジューラ

