

テーマ4：医療データ・医療LLM/LMMの利活用を促進する 医療データ基盤

◎中野 美由紀^{*1}, ◎田村 孝之^{*1}, 加藤 誠^{*2}, 天笠 俊之^{*2}, 吉川 正俊^{*3}, 宮崎 純^{*4},
曹 洋^{*4}, 福水 健次^{*5}, 合原 一幸^{*6}

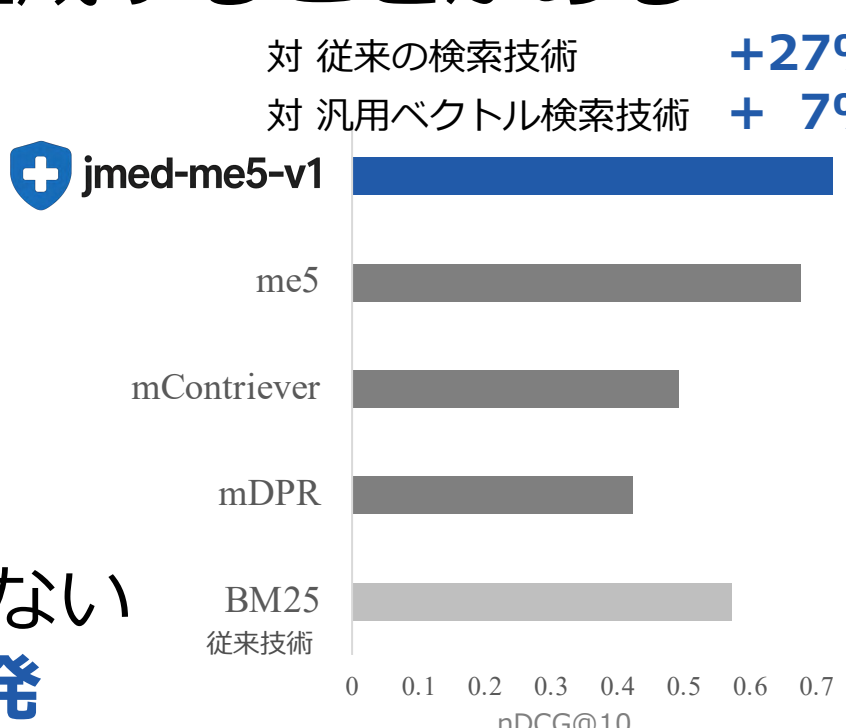
^{*1})情報・システム研究機構, ^{*2})筑波大学, ^{*3})大阪成蹊大学, ^{*4})東京科学大学,
^{*5})情報・システム研究機構 統計数理研究所, ^{*6})東京大学

医療データ基盤の想定利用シーン（デモ）

- 医療データ基盤を検索し、医療特化型モデルを取得
- モデルの出力を評価（ベクトルDBを検索）し、ハルシネーションの有無を検査
- モデルの来歴を用い、ベースとなった汎用モデルおよび学習データを分析

ハルシネーション対策（筑波大学）

- 背景：LLMは不正確な回答（ハルシネーション）を生成することがある
- 目的：LLMの回答と信頼できる情報源を照合し、回答の正誤を検証できるシステムを構築
- チャレンジ：
 - ①従来の検索技術では単語が一致しないと検索できない
→**ベクトル検索技術を照合に利用**
 - ②既存ベクトル検索モデルは日本語や医療文書に特化していない
→**日本語医療用ベクトル検索モデル jmed-me5-v1 を開発**

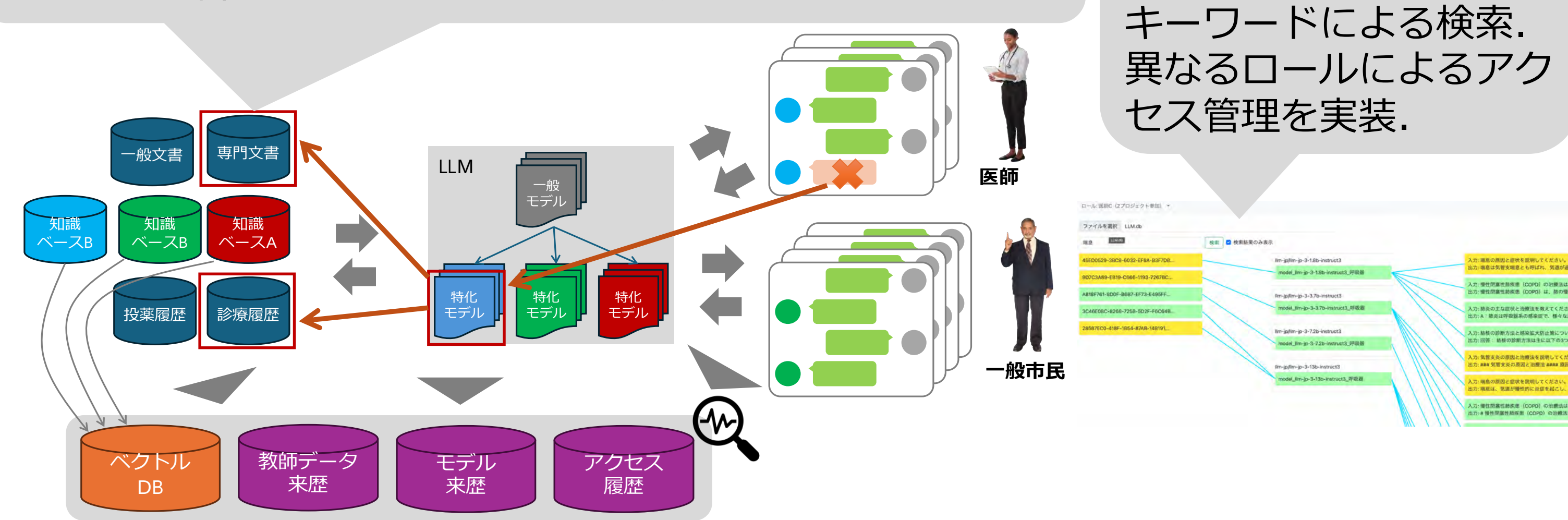


モデルおよびデータの来歴管理（筑波大学・大阪成蹊大学）

- 問題：医療用LLM/LMMの異なるモデルに対して、1) 学習データセット、2) モデルを使って得られた結果出力及び対応する入力プロンプトが急速に増加する一方で、LLM/LMMにおいてハルシネーションやエラーは本質的に避けることが困難。
- 動機：何らかの問題が生じたときに、関連するモデルやデータを瞬時に特定したい。
 - ✓ アドホックな管理では、特定に膨大なコストを要する上、最悪データが特定不可能な場合もある。
- アプローチ：モデル・データの来歴情報をシステムレベルで管理、アクセス可能にする。

教師データ、LLM/LMMモデル、入出力（アクセス履歴）をDB管理。問題の生じたデータを出発点として、データを探索。来歴情報の管理にはグラフモデルを利用。

プロトタイプUI
データ、モデル、入出力の依存関係を可視化。キーワードによる検索。異なるロールによるアクセス管理を実装。



LLMのためのベクトル索引（東京科学大学）

LLMの問題点

- ✓ 未学習知識に対して答えられない
 - ✓ ドメイン（応用）の固有情報など
- ✓ 学習済にも関わらず不正確な回答
 - ✓ ハルシネーション（幻覚）

ベクトルデータベース

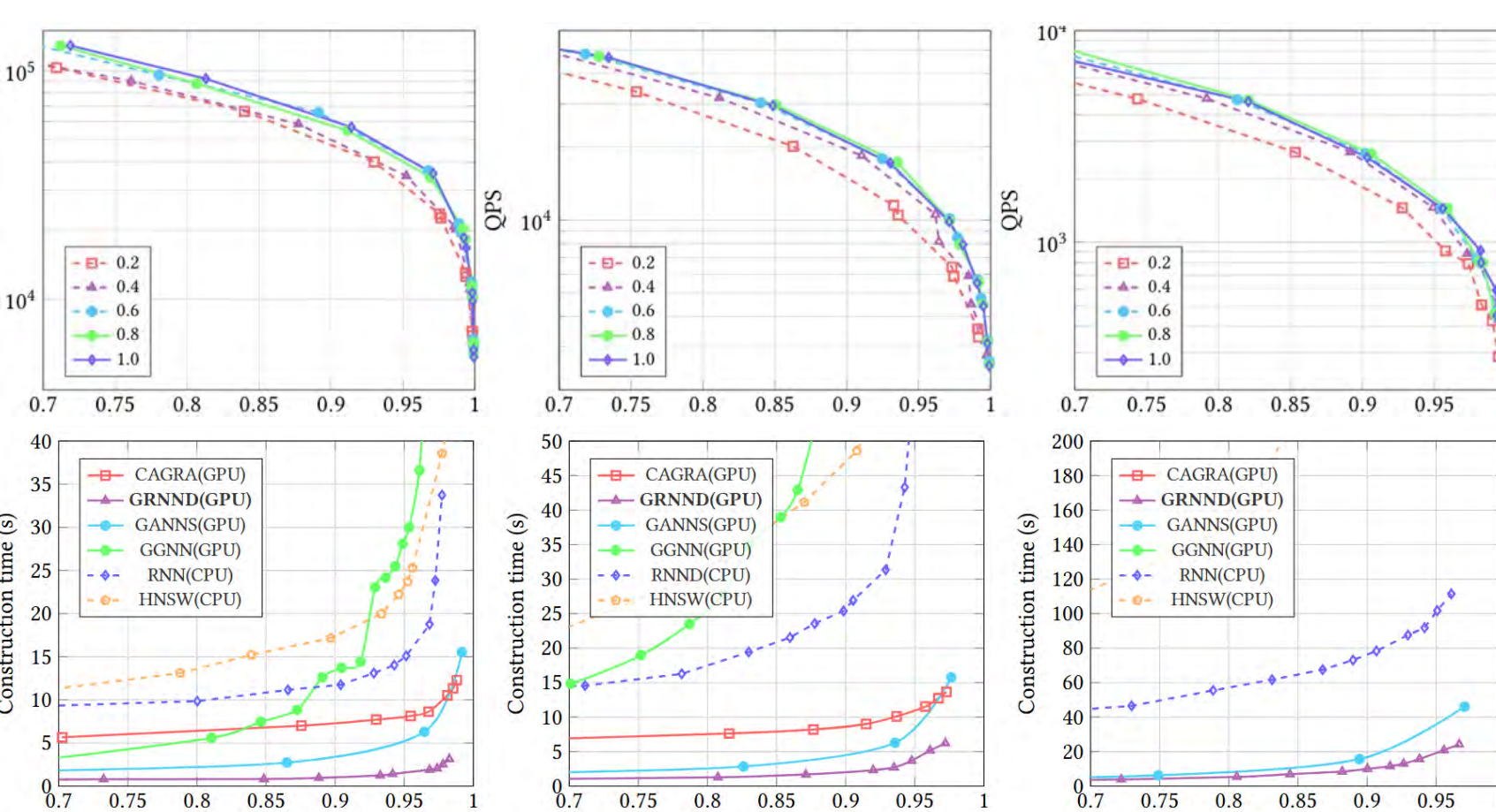
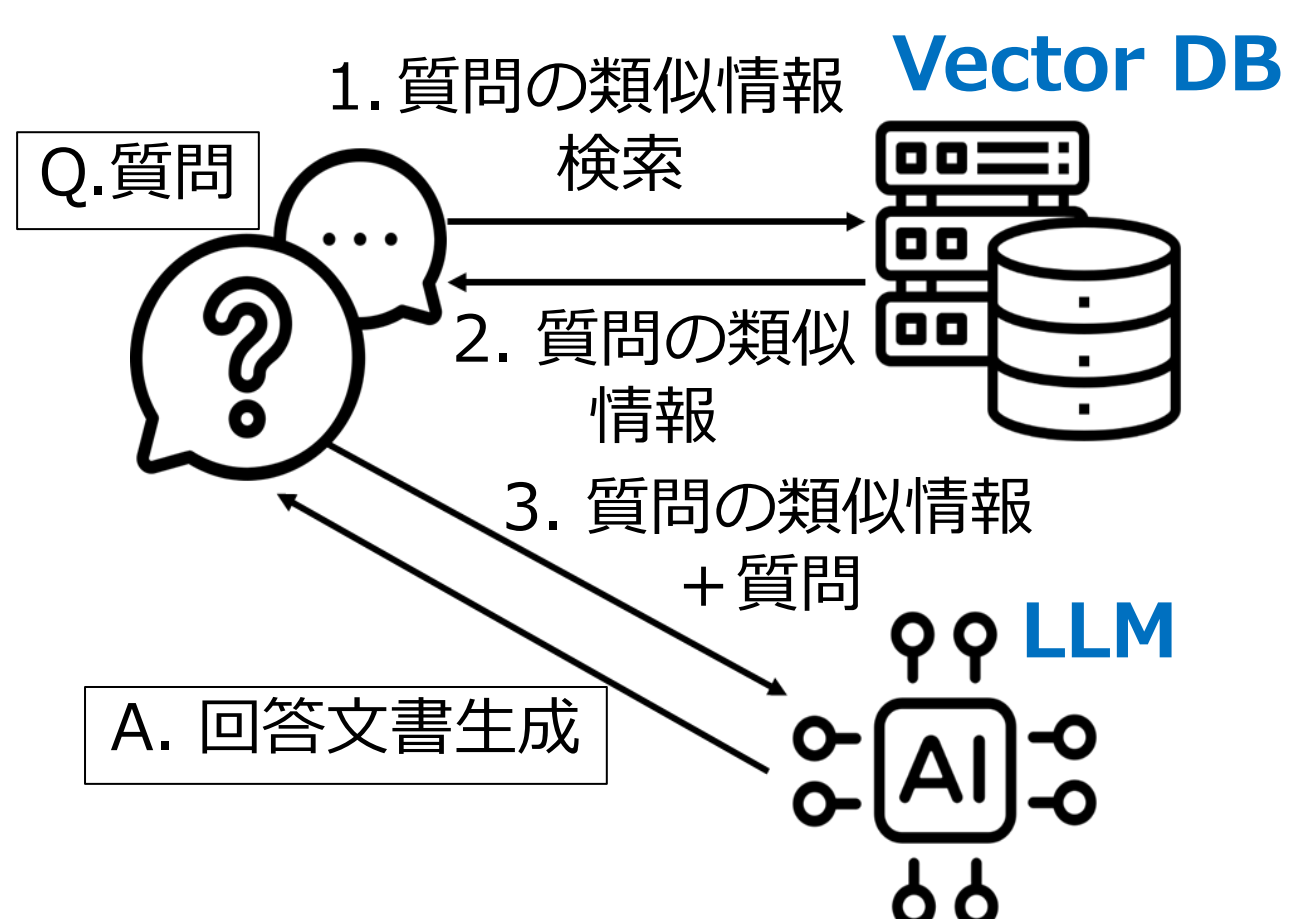
- ✓ LLMと別に文書の埋込み(ベクトル)を保存
 - ✓ LLM未学習情報の補完(Retrieval Augmented Generation)、ファクトチェック
- ✓ 重要機能は**類似ベクトルの検索**

ベクトルデータベースの課題

- ✓ 類似ベクトル検索用の索引構築時間
 - ✓ 例) 10億ベクトルで3ヶ月

チャレンジ

- ✓ GPUを利用した高速な索引構築
 - ✓ **GRND**: RNN-Descentを元にGPUに特化したアルゴリズム
- ✓ 結果
 - ✓ 既存手法より**2.2~6.9倍の構築速度**
 - ✓ 検索速度・精度は同等

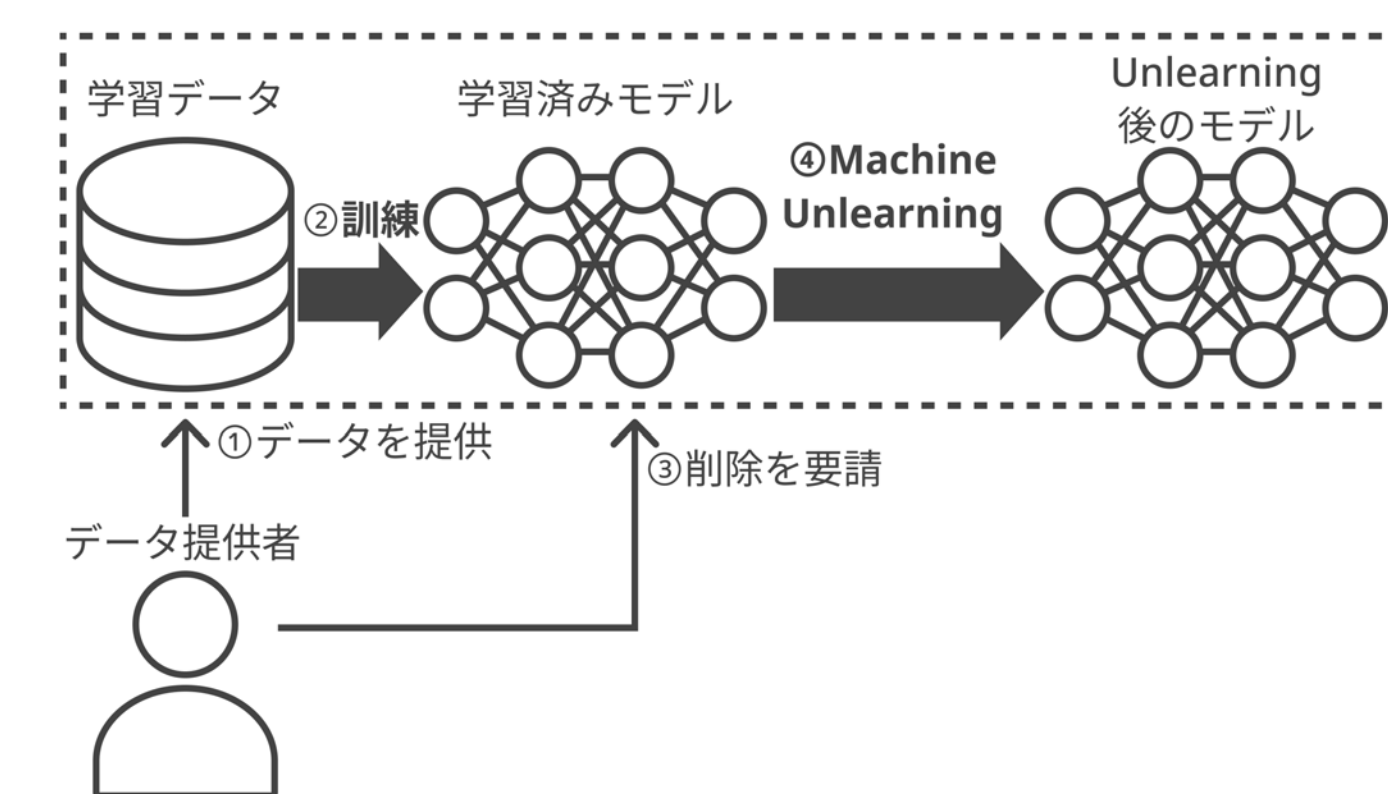


LLM/LMMの理解・制御に向けた取組み

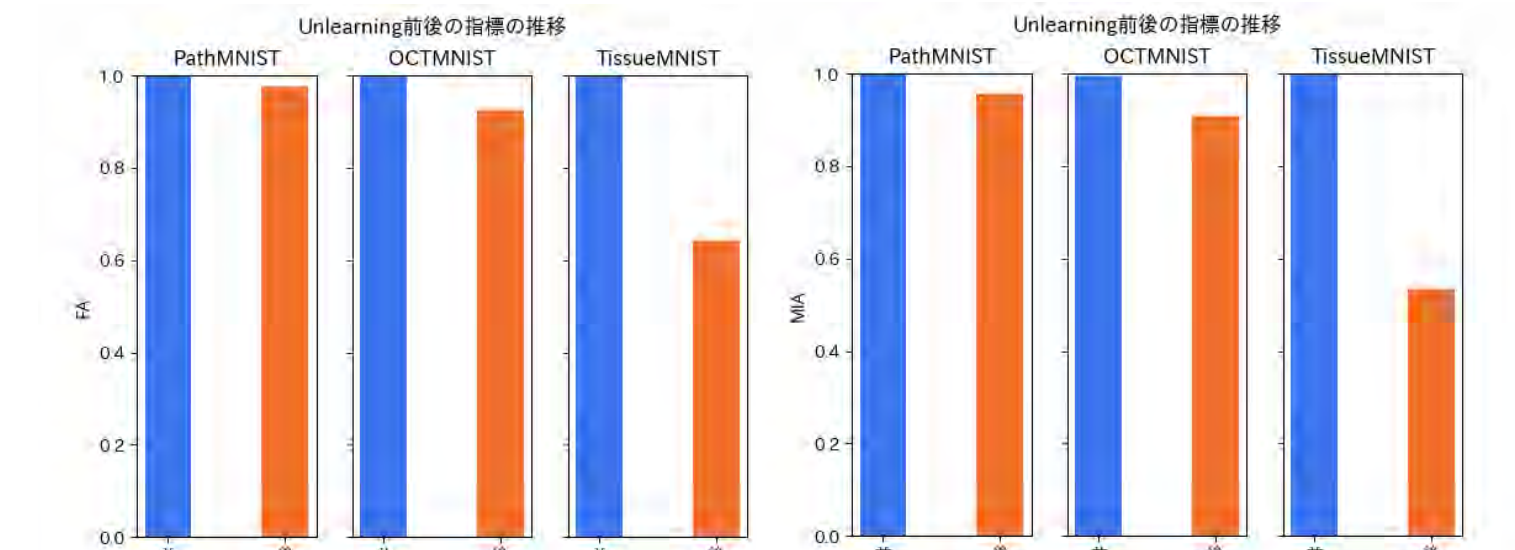
医療LMMにおけるUnlearning技術（東京科学大学）

- 研究目的：医療LMMにおいて、訓練データに不適切な情報、例えば、患者の個人情報や誤りを含むデータ等をモデルから「忘れさせる」ことができる**アンラーニング技術**を開発し、その効果を測定するための指標を構築すること。

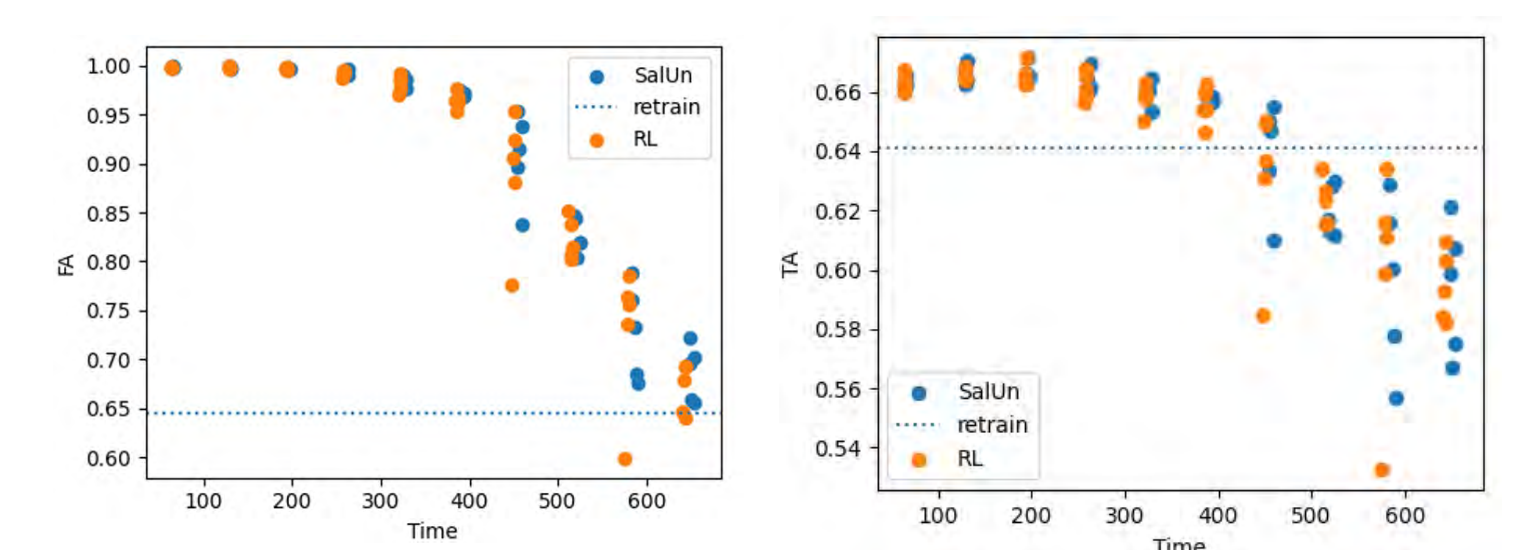
Unlearnの流れ



- 実験結果1: データセットによって指標の振る舞いが異なる



- 実験結果2: Unlearnの有効性が高まるにつれ、Unlearnしたモデルの正解率(TA)が低下する



Unlearn手法

- ランダムなラベルの貼り直し [CVPR20]
- Saliency-based Unlearning [ICLR24]

Unlearn指標

- 忘れさせたいデータでの正解率 (FA)
- メンバーシップ推論攻撃(MIA)

[CVPR20] Eternal sunshine of the spotless net: Selective forgetting in deep networks.
[ICLR24] SalUn: Empowering Machine Unlearning via Gradient-based Weight Saliency in Both Image Classification and Generation

医療基盤モデル構築のためのモデル融合進化技術の開発（統計数理研究所）

研究の目的：大規模基盤モデルを利用した医療データ

- 医療用LLM/LMM構築のためのデータ科学的課題の解決
- データ・モデルの多様性
 - **モデルの表現**：LLM/LMMの特徴・類似度を表現・比較する方法の開発
 - **モデル選択**：解決したい個別医療タスクに適した複数のモデルを選択する技術の開発
- タスク解決のための目的別モデル構築
 - 選択した複数の**モデルの融合・進化技術**の開発

研究結果：

LLMの地図構築法

- **提案したモデル類似度**に基づくモデル間の相互関係
- Hugging face（公開プラットフォーム）上のBert系モデル71個
- TriviaQAデータによるfine-tuningと類似度計算
- 性能との相関
 - QAタスク(TriviaQA)の正答率が近いモデルのクラスが存在

今後の展開：

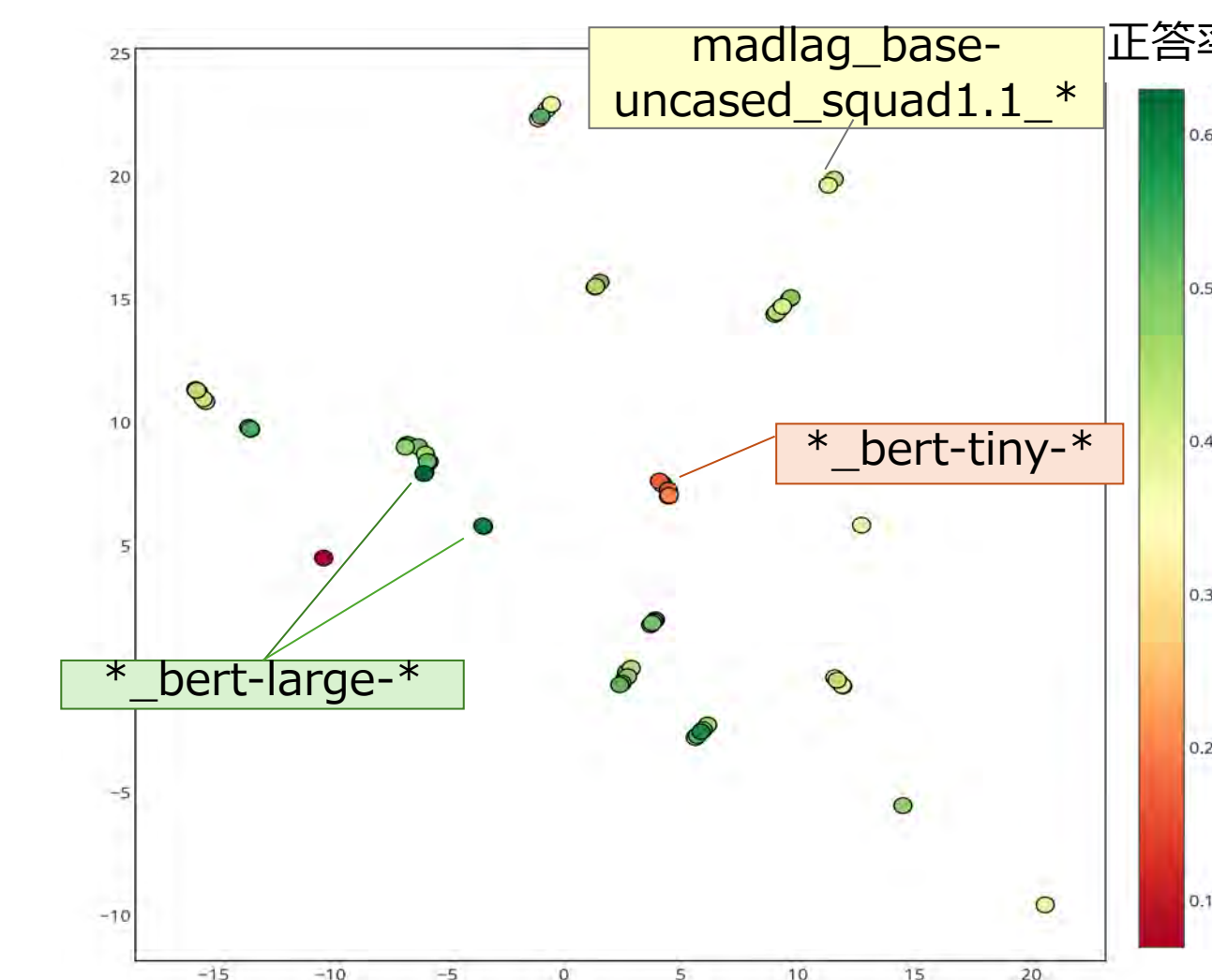
- 類似度を利用したモデル選択・融合法の開発
- 医療タスクに特化したLLM地図の構築

新しいモデル類似度の提案

Model a と Model b の類似度: S_{ab}

$$S_{ab} = \left\langle \begin{matrix} \text{類似度行列} \\ \text{類似度行列} \end{matrix} \right\rangle$$

LLMの地図（Bert系モデル）



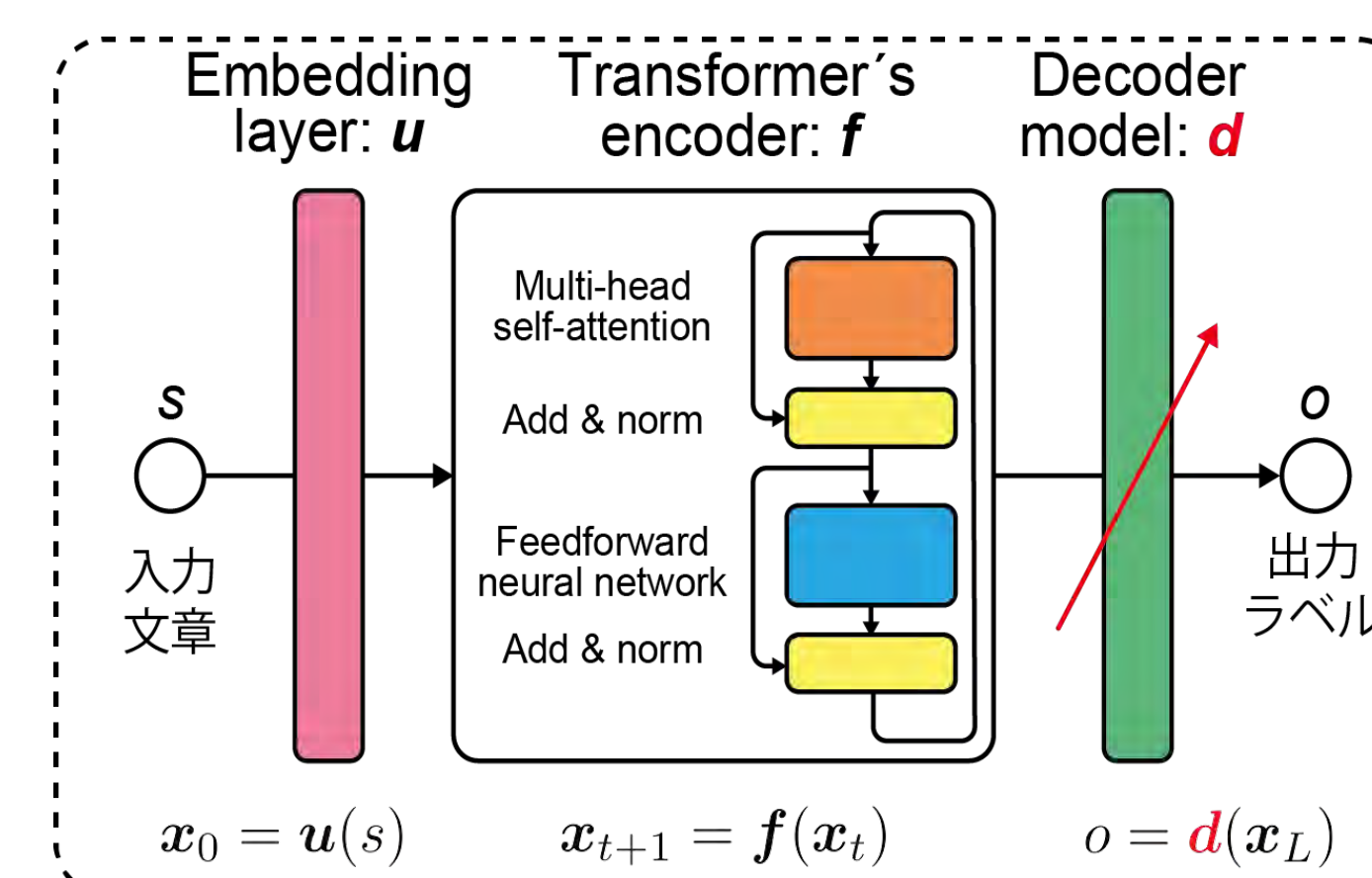
UMAPによる2次元表示。性能の高いクラスター、低いクラスターが存在。由来の似たモデルが集まる傾向

医療LLM/LMMの非線形ダイナミクス解析技術の開発（東京大学）

- リザーバー計算理論とエネルギー地形解析理論の統合で、LLM/LMMの数理構造・非線形ダイナミクスと脳の数理構造・非線形ダイナミクスの関係を考察することが可能な数理基礎技術を構築した。
- BERT/ALBERTが事前学習によって自然言語構造が封入された高次元リカレントニューラルネットワークであるリザーバーネットワークとして理解できることや、発達障がい圏固有の知性・考え方を生み出している神経動態にAI内部の信号処理ダイナミクスを近づけることでユニークな特性を持ったAIを開発できる可能性が見出された。

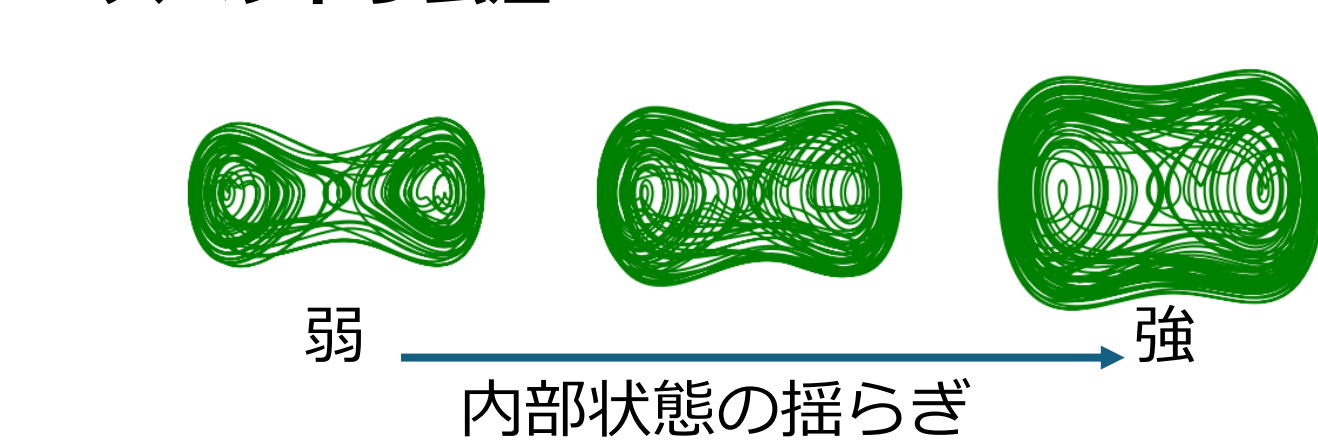
力学系としてのBERT/ALBERTモデル

Reservoir computingにおけるランダムなNNに代え、**自然言語**に特化されたNN (i.e., **"the reservoir"**)を使用



発達障がいの特性を生かしたAIの開発

自閉スペクトラム症 定型発達 注意欠陥多動性障害 (ADHD)



	秩序	カオスの縁	カオス
学習に必要なデータ量	小	中	大
汎用性	低い。	中程度	高い。 ひらめきも？

本展示に関し、開示すべきCOI関係にある企業などとして、

受託研究・共同研究費：ソフトバンク、LINEヤフー、日本電信電話