ABCI 3.0開発加速利用 （2025年度）成果概要（公開用）

課題名：偽情報対策特化型LLMの開発

実施時期：2025/4/16～2026/3/30
所属機関名：富士通株式会社
代表者氏名：小林健一

成果概要：NEDOのプロジェクト「経済安全保障重要技術育成プログラム／偽情報分析に係る技術の開発」の研究項目として、偽情報対策において実用的な性能を備え、かつ安全な特化型LLMを実現するため、既存のローカルLLMに対して報道領域とソーシャルメディア領域の言語能力の向上と偽情報領域の問題解決能力の向上を目的とする継続事前学習とファインチューニングを実施した。偽情報対策において重要な3つのタスクにおいて、ローカルLLMの利点である学習容易性を活かした継続事前学習とファインチューニングの併用により、最新のクラウドLLMの性能を凌駕した。

成果のポイント：

特化型LLM開発：安全性を重視しての閉環境動作が可能な既存ローカルLLMを対象にCPT＋SFTで偽情報対策向けに特化。

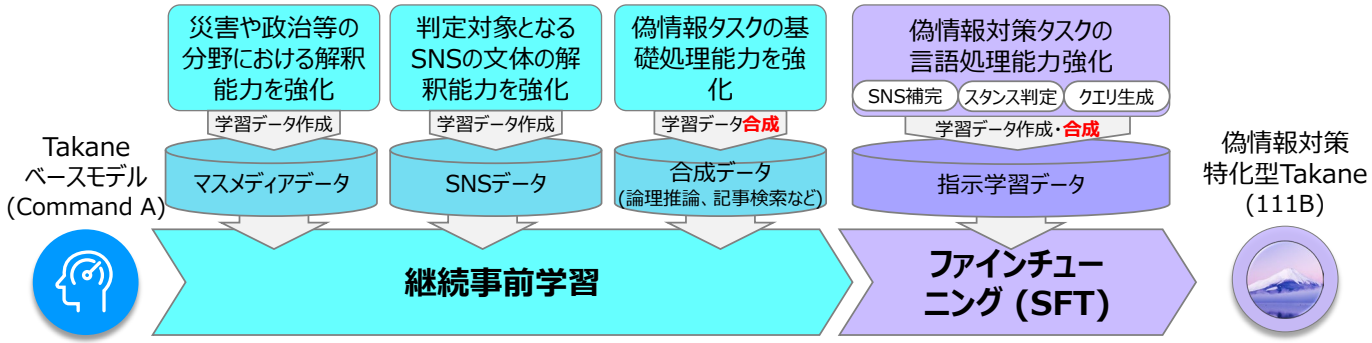
偽情報対策重要3タスク：偽情報対策の中核となる3タスク（SNS補完・スタンス判定・クエリ生成）の解決能力を強化。

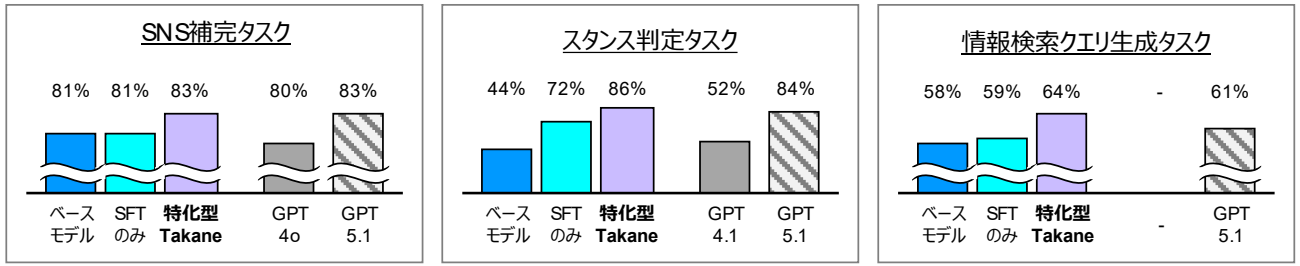
CPT（報道・SNS・偽情報）：高品質文書の選別・フィルタリングを行い、報道文章を起点に領域強化の合成データも活用して知識・文体理解能力を強化。

SFT（重要3タスク）：3つのタスクそれぞれについて指示データでSFTを実施し、タスク解決能力を強化。

評価結果：3タスクでベースモデルから性能向上を確認し、CPT＋SFTにより（クラウドLLMの一例として）GPT-4/5の精度超えを達成。

[謝辞]
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）の委託業務（JPNP22007）の結果得られたものです





成果についてより詳細な情報を提供しているWebページ、発表論文などの情報：

・論文：清水ら「LLMの特化学習による偽情報判定タスクの精度向上について」（DICOMO2026シンポジウム，2026年6月発表予定）
・プロジェクト採択時(2024年7月)のNEDOのプレスリリース: https://www.nedo.go.jp/news/press/AA5_101763.html
・同、富士通のプレスリリース: https://pr.fujitsu.com/jp/news/2024/07/19.html
・プロジェクト開始時(2024年10月)の富士通のプレスリリース: https://pr.fujitsu.com/jp/news/2024/10/16.html