

課題名：最先端AIモデルの研究開発

実施時期：2025年度  
 所属機関名：東京科学大学  
 代表者氏名：岡崎 直観

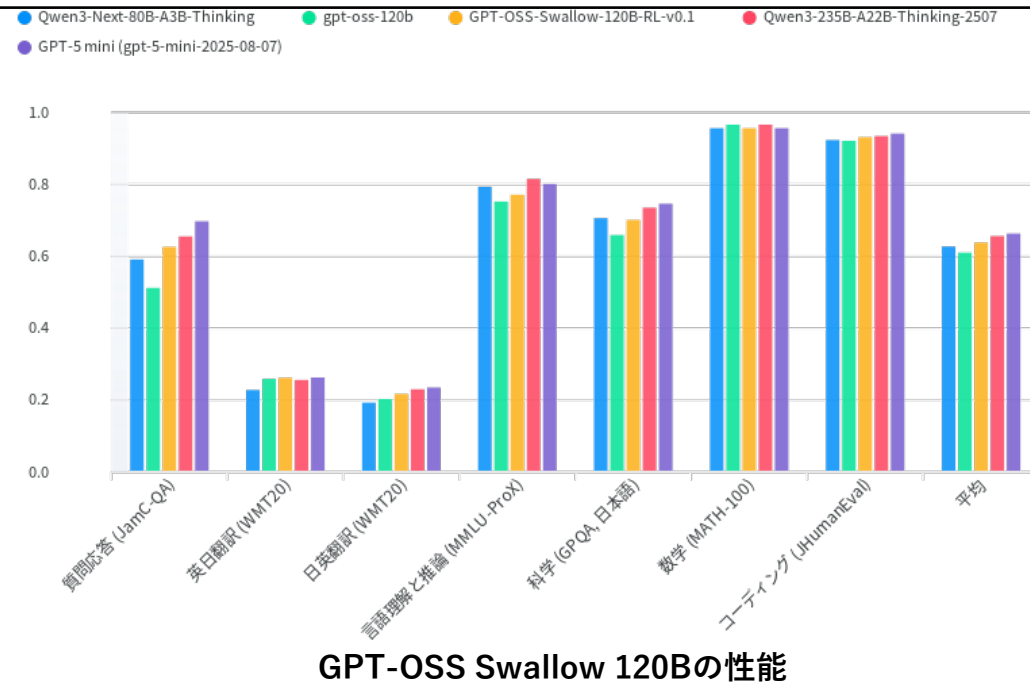
成果概要：大規模言語モデルSwallow (Swallow LLM) シリーズとして、Qwen3-Swallow 8B, 30B-A3B, 32BおよびGPT-OSS Swallow 20B, 120Bを構築し、Apache 2.0ライセンスでモデルを公開した。これらのモデルを深い推論に対応させるため、Swallowプロジェクトで用いる評価フレームワークの刷新、推論過程付きの事後学習データの合成、検証可能な報酬による強化学習を導入した。その他、マルチモーダル基盤モデル、LLMの安全性を高めるための先端的な研究を実施した。

成果のポイント：

- 深い推論に対応したLLMを構築するため、高難易度タスクを含む評価フレームワークであるswallow-evaluation-instructを開発した。
- 推論過程付きの事後学習データを構築した。
- 検証可能な報酬による強化学習を導入した。

Qwen3およびgpt-ossをベースモデルとして、以上の成果を投入して継続事前学習・事後学習を行い、Qwen3 SwallowおよびGPT-OSS Swallowを構築・公開した(右図)。GPT-OSS Swallow 120B RLはオープンなLLM(120B以下)の中で日本語・英語タスクの両方で最高性能を達成した(右図)。

その他、LLMの長文処理能力の改善、LLM生成文の自動検出、視覚言語モデルのための学習データ・評価データの構築などの研究を進めた。



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：

Swallow LLMプロジェクトページ: <https://swallow-llm.github.io/index.ja.html>

Swallow LLMのHuggingFaceページ: <https://huggingface.co/tokyotech-llm>

東京科学大学情報理工学院・岡崎研究室の発表文献リスト: <https://www.nlp.c.titech.ac.jp/publications.ja.html>