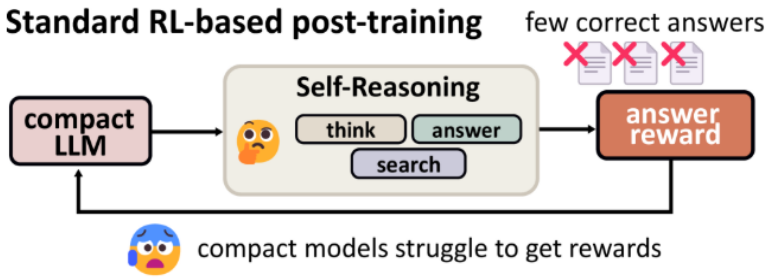
ABCI 3.0開発加速利用 （2025年度）成果概要（公開用）

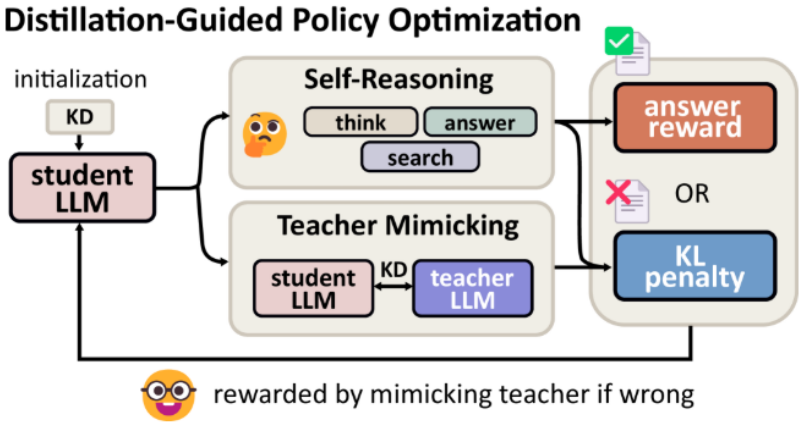
課題名：
大規模ベクトル検索を応用した基盤モデルの拡張と生涯学習に関する研究開発

実施時期：2025/06/24 ～ 2026/03/30
所属機関名：オムロン サイニックエックス株式会社
代表者氏名： 西村真衣

成果概要：
コンパクト言語モデル（0.5–1Bパラメータ）への Agentic RAGポストトレーニングを実現する手法 Distillation-Guided Policy Optimization（DGPO）を提案・実装し、ACL 2026 Main（Oral）に採択された。本研究ではABCI H200 141GB を活用し、知識蒸留を伴う強化学習の大規模実験を実施した。

成果のポイント：
強化学習（RL）はLLMの事後学習おいて有効な手法だが、超小規模言語モデル（0.5–1Bパラメータ）では初期性能の低さから報酬が得られにくく、学習崩壊が生じる。本研究では、教師モデルによる蒸留をコールドスタート初期化とRL中のガイダンスの両面に統合したDGPOを提案した。
探索失敗時は教師模倣を学習信号とし、探索成功時は自己推論を報酬化するという効果的な枠組みにより、0.5–1BモデルでのAgentic Searchを初めて実現した。さらに、蒸留単独では不可能な「学生モデルが教師モデルを超える」性能を達成した。





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：
https://github.com/omron-sinicx/dgpo
https://omron-sinicx.github.io/dgpo/
R.Kotoge,M.Nishimura,J.Ma, "Can Compact Language Models Search Like Agents? Distillation-Guided Policy Optimization for Preserving Agentic RAG Capabilities", In Proceedings of The 64th Annual Meeting of the Association for Computational Linguistics (ACL 2026) Main.