継続事前学習による
推論型大規模言語モデルの構築

岡崎 直観

東京科学大学 情報理工学院
okazaki@comp.isct.ac.jp
https://www.nlp.c.titech.ac.jp/

Institute of SCIENCE TOKYO

OKAZAKILAB

Swallowプロジェクト

日本語で高性能な大規模言語モデル

日本に関する知識が豊富なだけでなく、数学、コーディング、推論にも強い、汎用型の大規模言語モデルを志向

商用利用可能なライセンスで公開

できるだけ利用制限の少ないライセンスを採用し、構築したモデルをHuggingFace上で公開

アカデミックな研究開発

東京科学大学の岡崎研究室、横田研究室、産業技術総合研究所のメンバーを中心に開発

オープンな研究開発

高性能なモデルを構築するためのレシピ、訓練データ、実験結果を共有し、日本の人工知能研究・応用を促進

実績（2026年3月時点）

241万	132	55万	19
モデルのダウンロード	モデルの数	データセットのダウンロード	データセットの数

https://swallow-llm.github.io/index.ja.html

Swallowプロジェクトの歩み

Swallow (Llama 2) (2023/12)

- 🏆 オープンなLLMの中で日本語で最高性能
- 👉 語彙拡張（日本語の16kトークンを追加）
- 📄 言語処理学会で3件の発表（2件は優秀賞）
- 📄 COLM 2024で2件の発表

Swallow (Mistral/Mixtral) (2024/3)

- 🏆 オープンなLLMの中で日本語で最高性能
- 👉 Mixture of Expert (MoE) に対応

Llama 3 Swallow (2024/6)

- 👉 継続事前学習データの配合を改良

Llama 3.1 Swallow (2024/10)

- 🏆 オープンなLLMの中で日本語で最高性能
- 👉 継続事前学習データの品質を改善（学習データの品質フィルタリング、合成データ）

https://swallow-llm.github.io/index.ja.html

Llama 3.3 Swallow (2025/3)

- 🏆 オープンなLLMの中で日本語で最高性能
- 👉 指示チューニングデータの合成
- 📄 COLM 2025で1件の発表
- 👉 数学とコーディングの改善 (Swallow Code v1)
- 📄 ICLR 2026で1件の発表

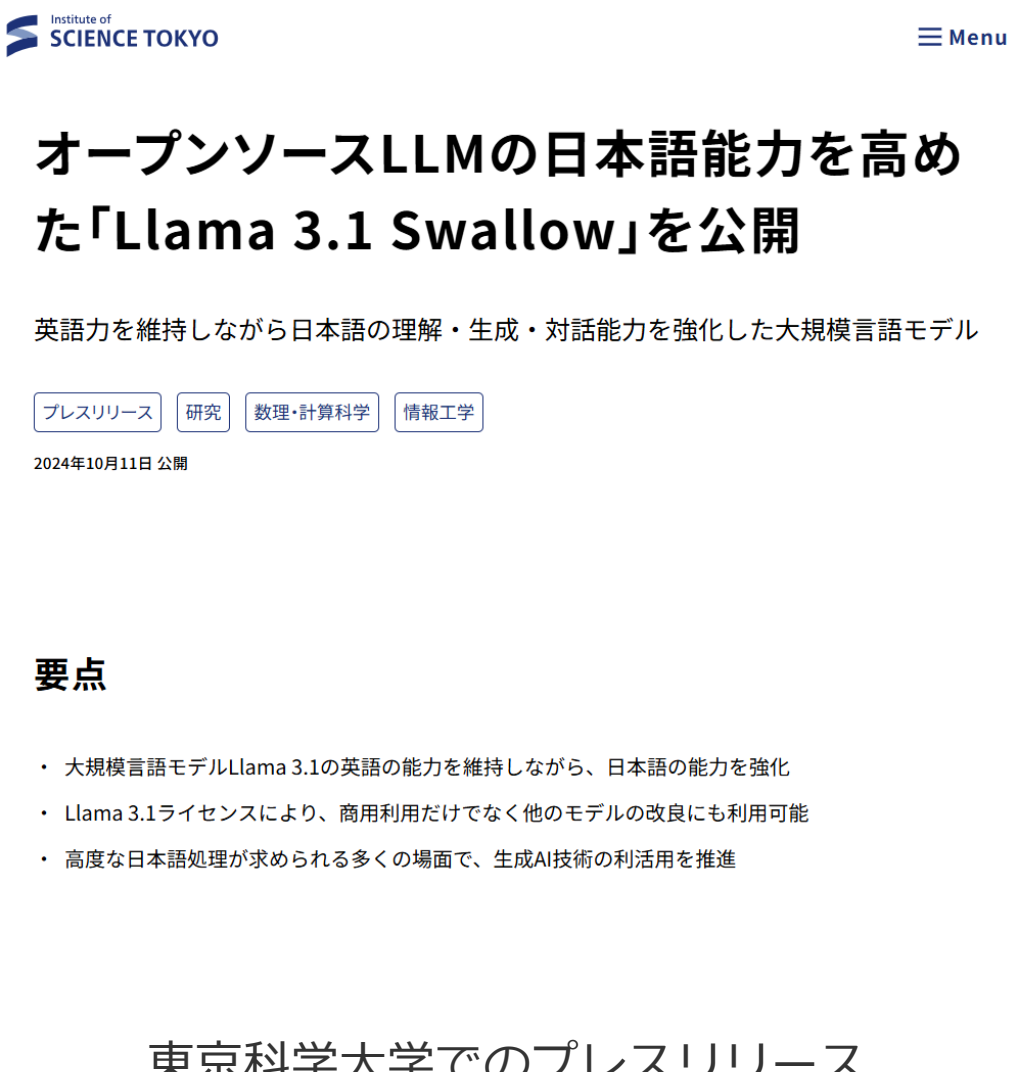
Gemma2-Llama Swallow (2025/5)

- 🏆 オープンなLLMの中で日本語で最高性能
- 👉 Llama 3.3 Swallowのレシピを採用
- 👉 TPU環境による学習

Qwen3 Swallow, GPT-OSS Swallow (2026/2)

- 🏆 オープンなLLMの中で日本語で最高性能
- 👉 Apache 2.0ライセンス
- 👉 推論型モデル向けの事後学習データ
- 👉 検証可能な報酬による強化学習 (RLVR)

Swallowに関するプレス発表



Institute of SCIENCE TOKYO

Menu

オープンソースLLMの日本語能力を高めた「Llama 3.1 Swallow」を公開

英語力を維持しながら日本語の理解・生成・対話能力を強化した大規模言語モデル

プレスリリース　研究　数理・計算科学　情報工学

2024年10月11日 公開

要点

- 大規模言語モデルLlama 3.1の英語の能力を維持しながら、日本語の能力を強化
- Llama 3.1ライセンスにより、商用利用だけでなく他のモデルの改良にも利用可能
- 高度な日本語処理が求められる多くの場面で、生成AI技術の利活用を推進

東京科学大学でのプレスリリース

オープンソースLLMの日本語能力を高めた「Llama 3.1 Swallow」を公開（2024年10月11日） https://www.isct.ac.jp/ja/news/g3j45hj4otpa

NVIDIA DEVELOPER ホーム ブログ フォーラム ドキュメント ダウンロード トレーニング

技術ブログ　ブログ記事を探す　絞り込む

Generative AI

先進的なソブリン AI モデルが、日本のイノベーションとチャンスを解き放つ

2024 年 10 月 9 日　+4 Like

By Chintan Patel, Mana Murakami and Naoyuki Yura

ソブリン AI モデルは、特定の文化的や言語的ニュアンスに合わせて調整されているため、文脈を理解し適切な応答を生成する上でより効果的です。さらに、これらのモデルは地域のイノベーションを支援し、各国がそれぞれのニーズや優先事項に沿った AI 技術を開発することを可能にします。

その結果、世界各国から支持されるソブリン AI モデルを開発しようという動きが高まっています。実際に、各国政府は、研究者や企業が自国民のニーズに特化した AI システムを構築できるようにするための構想を開始し、計算インフラに予算を割り当てています。

最も先進的な日本語 AI モデル

このたび、東京科学大学 (旧・東京工業大学) と産業技術総合研究所 (AIST)は、Llama 3.1 をベースに、日本特有の言語的/文化的ニーズによりよく応えるように設計された独自のソブリン AI モデル「Llama 3.1 Swallow」を共同開発しました。

研究チームでは、CommonCrawl から配布されているアーカイブ全量から、日本語のテキストを独自に抽出/精錬した Swallow Corpus Version 2 という日本語ウェブコーパスを構築し、NVIDIA NeMo でも利用可能な Megatron-LM を使用してモデル学習を行いました。この新しいコーパスは、前バージョンの Swallow model に使用された Swallow Corpus Version 1 よりも約 4 倍大きく、より包括的で文化的に適切な AI 能力を日本向けに強化しています。

最終的な学習データは、ウィキペディアのデータに数学やコーディングなどのコンテンツを混ぜて、約 2,000 億トークンで構成され、モデルの継続事前学習に使用されました。

NVIDIAのウェブサイトでSwallowを試せる

先進的なソブリン AI モデルが、日本のイノベーションとチャンスを解き放つ (2024年10月9日) https://developer.nvidia.com/ja-jp/blog/advanced-sovereign-ai-model-unlocks-innovation-and-opportunities-for-japanese-citizens/

Swallowの開発で活用している計算資源

ABCI 3.0 (産総研)

- 2025年1月20日サービス開始
- NVIDIA H200 SXM5 141GB x 8 x 766
- 半精度演算のピーク性能は6.2EFLOPS
  - ABCI 2.0から7～13倍の性能向上
- 時間単価（標準利用の場合）
  - バッチ: 3300円/時間・ノード
  - 予約: 4950円/時間・ノード

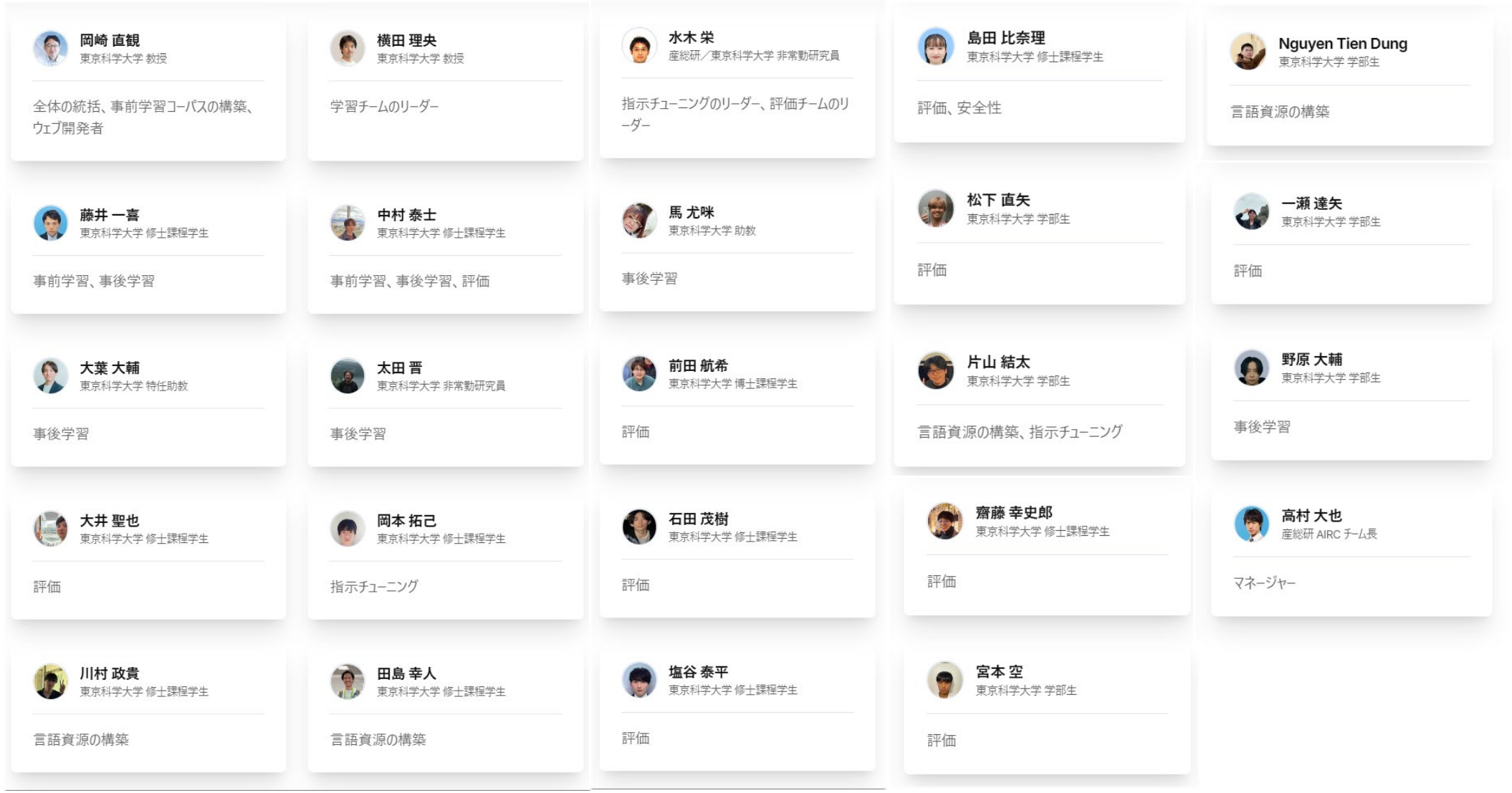
https://www.aist.go.jp/aist_j/news/pr20241010.html

TSUBAME 4.0 (東京科学大学)

- 2024年4月1日サービス開始
- NVIDIA H100 SXM5 94GB × 4 x 240
- 半精度演算のピーク性能は0.95EFLOPS
  - TSUBAME 3.0から5.5～20倍の性能向上
- 時間単価（学外・成果非公開の場合）
  - バッチ: 1100円/時間・ノード
  - 予約: 時期により単価が1.25～10倍で変動

https://www.titech.ac.jp/news/2024/069378

Swallowの開発チーム

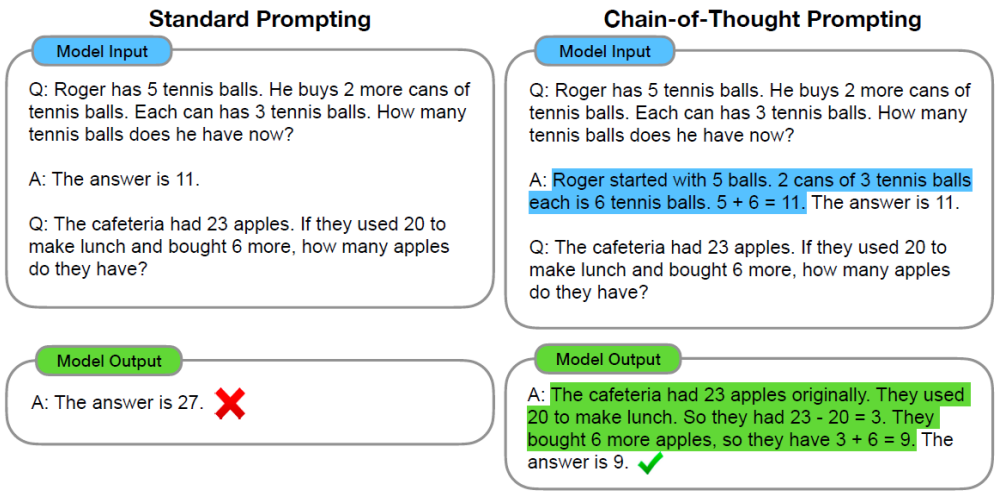


https://swallow-llm.github.io/index.ja.html

推論型のLLM

プロンプトで説明をするように誘導（従来）

- Few-shot事例や"Let's think step by step"などの指示をプロンプトに挿入することで、LLMに説明生成を促すと（思考の連鎖）、タスクの性能を高めた（ChatGPT登場前から知られていた）
- LLMに自問自答を繰り替えさせることで性能を改善する手法（自己改善）なども提案された



思考の連鎖 (CoT; Chain of thought)[1,2]

思考過程を生成するように学習（推論型）

- 事後学習（中間学習）や強化学習で推論過程（思考や計画、探索）を生成させてから回答をまとめるように学習したモデル
- 推論過程には与えられた問題の整理、解法の計画、試行錯誤（"Wait!"などのaha moment）、結論の取りまとめなどが含まれる
- （学習時の計算予算ではなく）推論時の計算予算を増やすことでLLMの性能を高める**推論時スケーリング**に対応
- LLMを汎用的な**問題解決エージェント**に近づける重要な技術
- OpenAI o1やDeepSeek-R1などが有名

[1] J Wei, X Wang, D Schuurmans, M Bosma, et. al. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *NeurIPS*.
[2] T Kojima, S S Gu, M Reid, Y Matsuo, Y Iwasawa. 2022. Large Language Models are Zero-Shot Reasoners. *NeurIPS*.

推論型の大規模言語モデルの生成例（MATH-100[2]の問題を解く例）

以下の数学の問題を、わかりやすく、論理的に解いてください。
出力の最後の行は、次の形式にしてください。

回答: $¥boxed{{ANSWER}}$

`ANSWER`には、問題の答えに対する最終的な数式または数値が入ります。

ステップバイステップで考えてから回答してください。

$f(x)=¥frac{2x}{x^2-5x-14}$ のグラフには、垂直漸近線 $x=a$ と $x=b$、水平漸近線 $y=c$ があります。$a+b+c$ を求めなさい。

プロンプトと問題

GPT-OSS 120B[1]

We need to find vertical asymptotes: denominator zero where numerator not zero. f(x)=2x/(x^2 -5x -14). Factor denominator: x^2 -5x -14 = (x-7)(x+2). So （…省略… ）
Then a+b+c = 7 + (-2) + 0 = 5. Answer 5.

Need to present step-by-step reasoning in Japanese. Output final line: 回答: $¥boxed{5}$.

Let's produce solution.

思考過程 (reasoning_content)

解答手順

1. **関数の分母を因数分解する**

$$x^2 - 5x - 14 = (x-7)(x+2)$$

2. **垂直漸近線（分母が 0 になる点）** 分母が 0 になるのは

$$x-7=0 \Rightarrow x=7, \qquad x+2=0 \Rightarrow x=-2$$

ただし、分子 $2x$ が 0 になる点は $x=0$ だけで、$x=7, -2$ では分子は 0 ではないので、両点とも垂直漸近線になる。

$$a=7,\ b=-2$$

3. **水平漸近線（$x \to \pm\infty$ のときの挙動）** 分子の次数は 1、分母の次数は 2 で、分子の次数が分母より低いので

$$\lim_{x\to\pm\infty} f(x) = 0$$

したがって水平漸近線は $y=0$。

$$c=0$$

4. **求める和**

$$a+b+c = 7+(-2)+0 = 5$$

最終結果

$$\boxed{5}$$

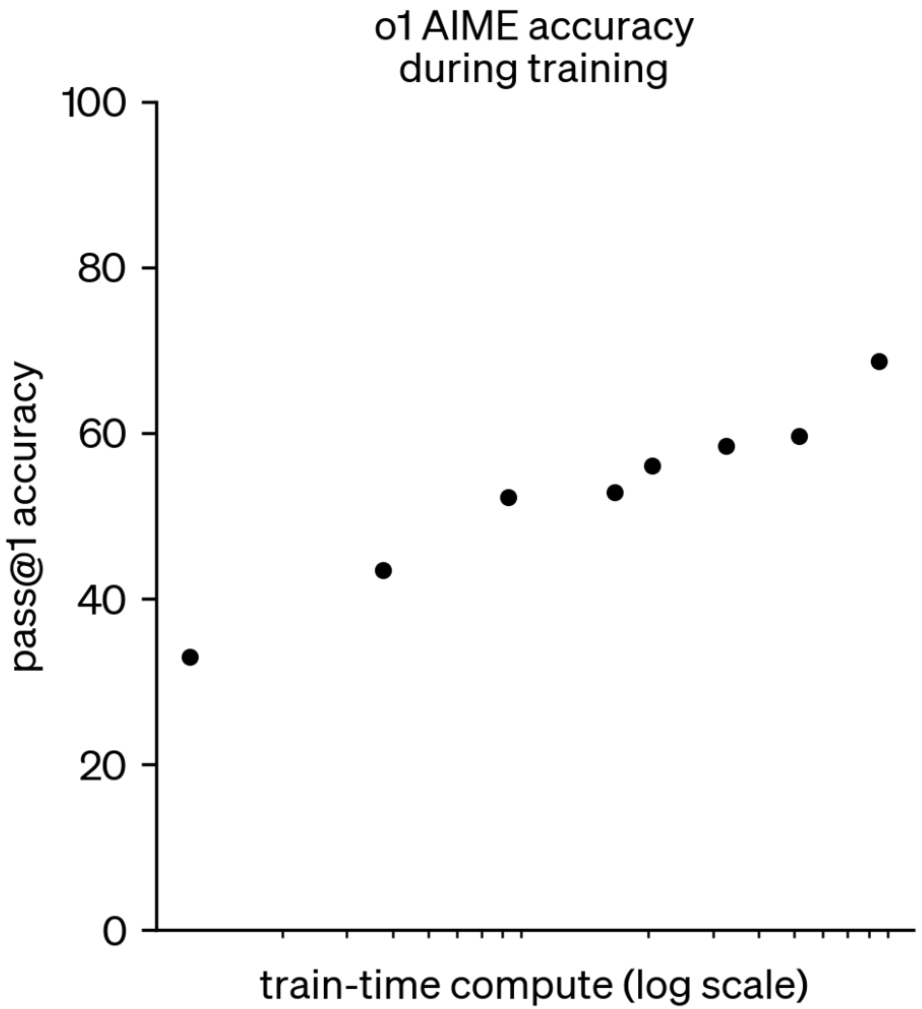
回答: $\boxed{5}$ ← 回答: $¥boxed{5}$

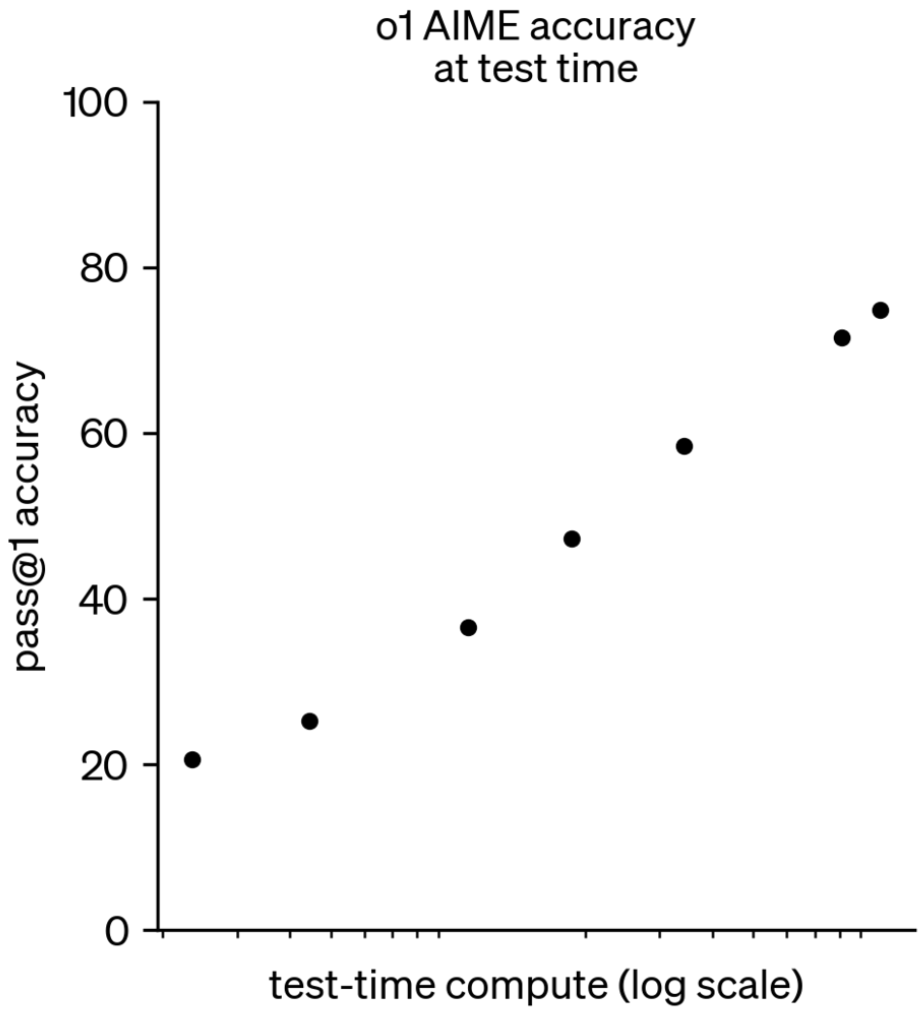
解答 (content)

Math-Verify[3]ライブラリで数式処理をして正解と比較

[1] https://huggingface.co/openai/gpt-oss-120b
[2] G Son, J Hong, H Ko, J Thorne. 2025. Linguistic Generalizability of Test-Time Scaling in Mathematical Reasoning. *ACL*.
[3] https://github.com/huggingface/Math-Verify (Lightevalに内蔵されている)

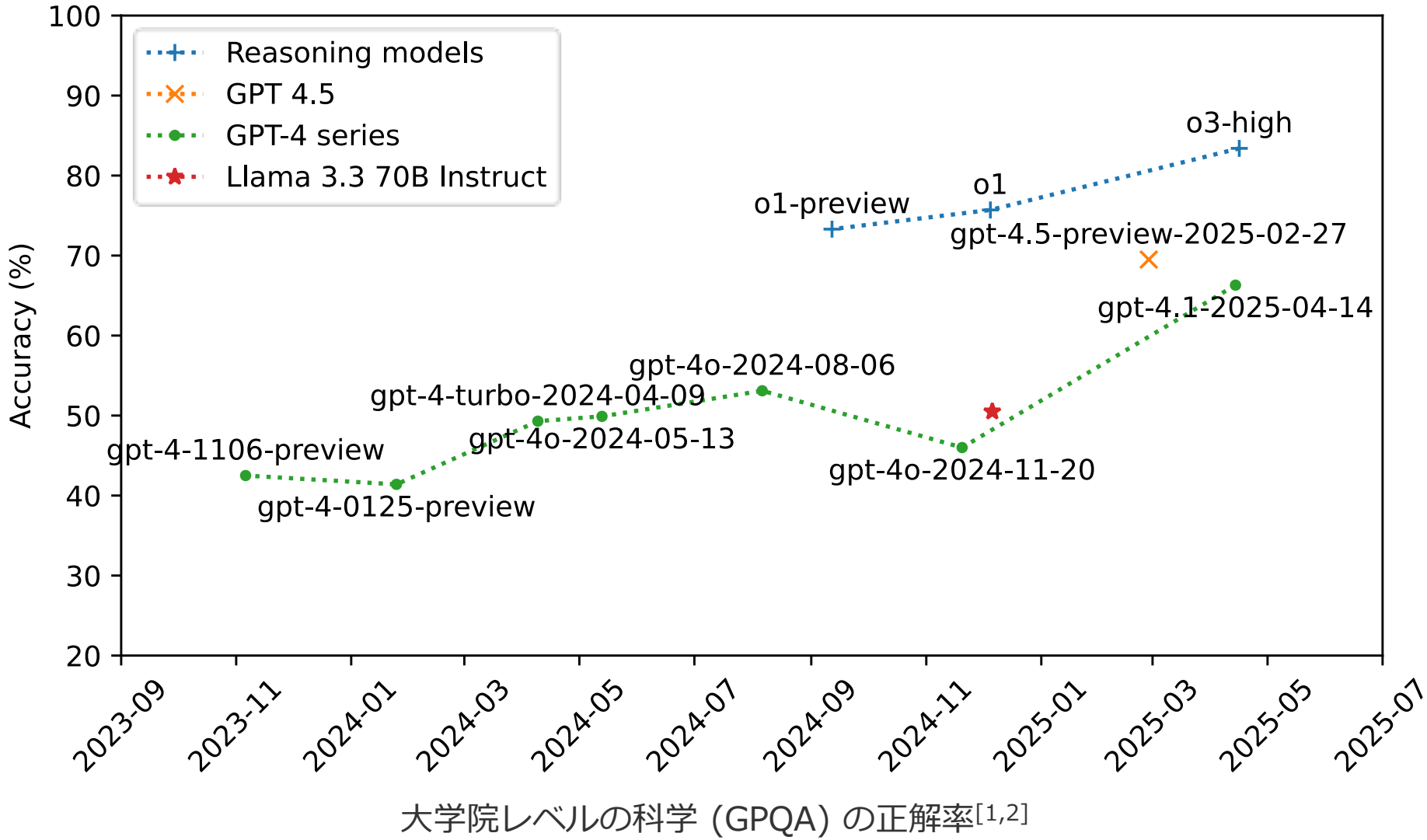
推論時スケーリング





https://openai.com/ja-JP/index/learning-to-reason-with-llms/

推論時スケーリングによるLLMの性能向上



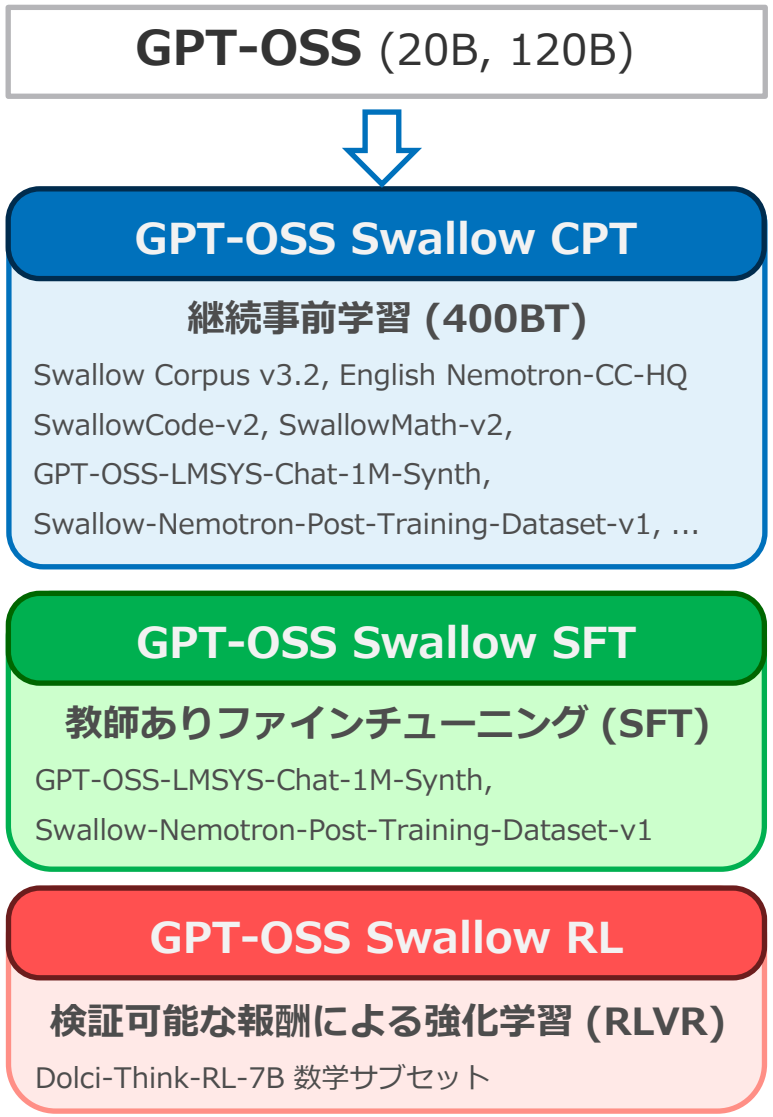
[1] D Rein, B L Hou, A C Stickland, J Petty, R Y Pang, J Dirani, J Michael, S R Bowman. 2024. GPQA: A Graduate-Level Google-Proof Q&A Benchmark. *Conference on Language Modeling (COLM)*.
[2] OpenAIが公開しているsimple-evalsとLlama 3.3 70B Instructのリリースノートを基にグラフ化

GPT-OSS SwallowとQwen3 Swallowの構築レシピ

GPT-OSS SwallowとQwen3 Swallowの構築方針

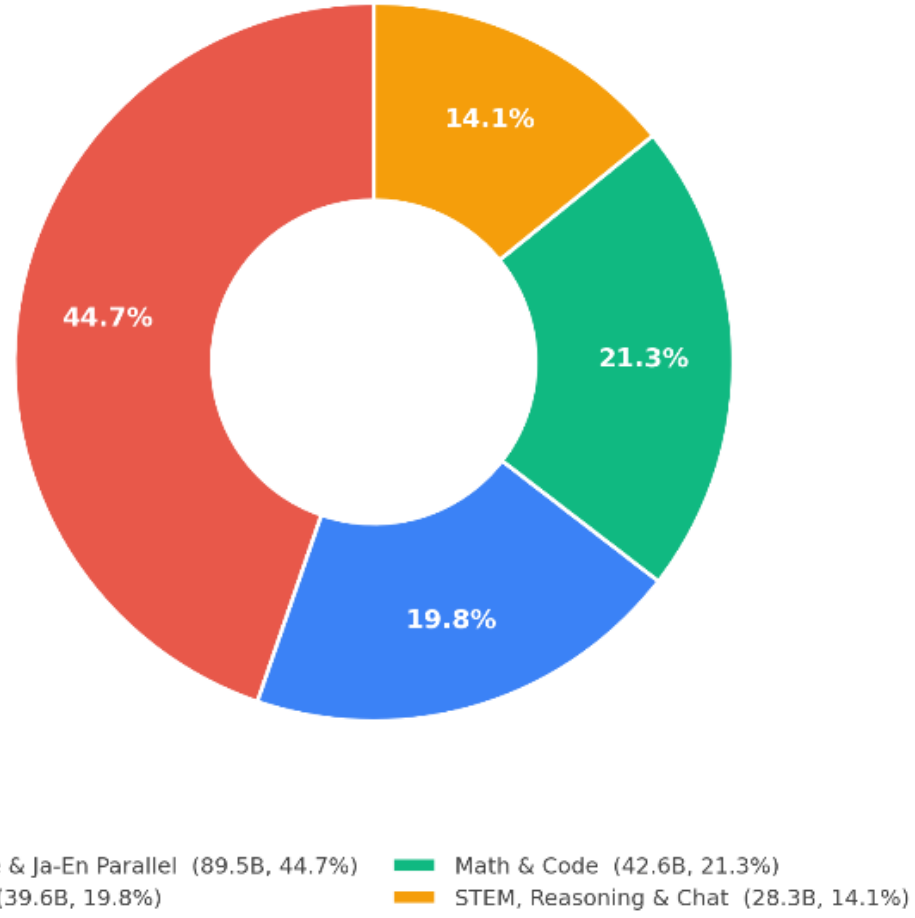
- 継続事前学習（CPT）、教師ありファインチューニング（SFT）、強化学習（RL）をGPT-OSS（20B, 120B）およびQwen3 (8B Base, 30B-A3B, 32B) に順に適用
  - GPT-OSSおよびQwen3は推論型であるが、継続事前学習を単に適用すると、**非推論型モデルに戻り、大幅な性能低下を招く**
- gpt-oss-120bで合成した推論過程付きデータを学習に活用
- 検証可能な報酬付の強化学習（RLVR; reinforcement learning with verifiable rewards）で推論能力をさらに強化

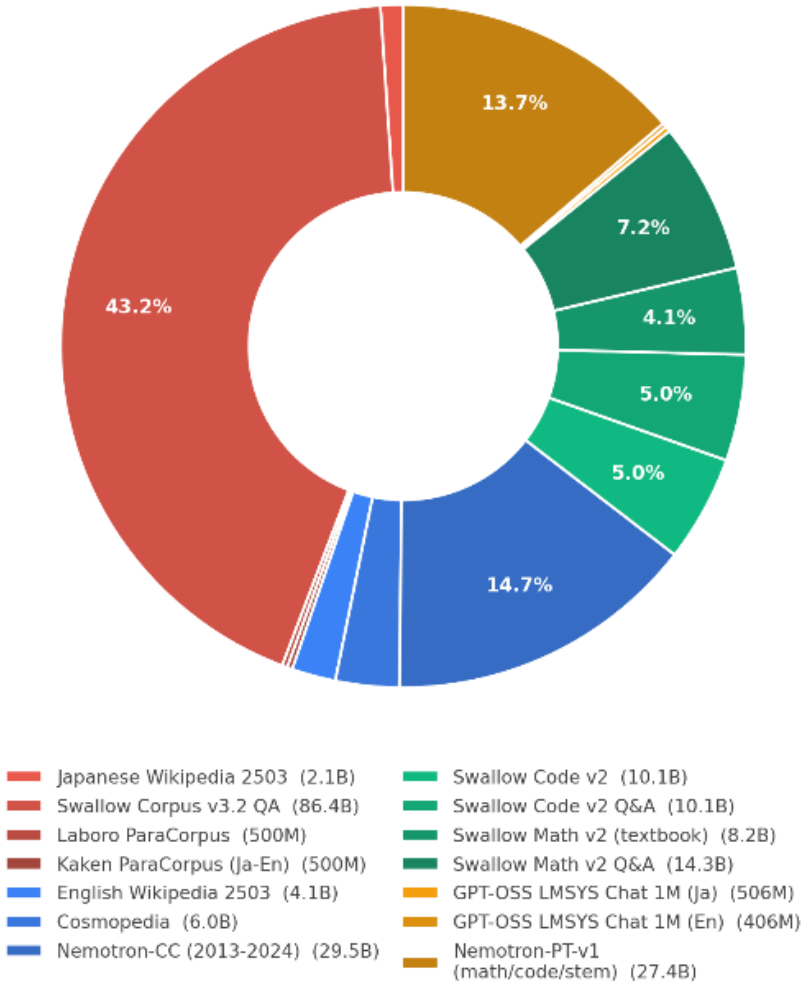
ステージ	学習内容
CPT	日本語と英語のテキスト、日本語の質問応答ペア、数学、コーディング、推論過程付きの指示と応答のペア
SFT	汎用対話およびSTEMにおける推論過程付きの指示と応答のペア
RL	数学設問の正誤（検証可能な報酬）



継続事前学習 (CPT)

日本語のウェブテキストを45%（日本固有の知識を身に付けさせるため）、英語のウェブテキストを20%（英語の忘却を防ぐため）、数学とコーディングを20%、推論過程付きの指示と応答のペアを15%配合





Swallow Corpus (Okazaki+ 2024) Version 3.2

Step 1	**WARCファイルのダウンロード**	Common Crawlと国語研日本語ウェブコーパスの全量
Step 2	**迅速な日本語判定**	日本語ではないと思われるウェブページ（約95%）を除去
Step 3	**HTMLからのテキスト抽出**	Trafilaturaを用いてHTMLからテキストを抽出
Step 4	**精密な日本語判定**	fastText (lid.176.bin) を用いてテキストの言語を検出

生コーパス: 11,627,589,798ページ, 17,348,032,012,899文字

Step 5	**MinHashによる重複除去**	マルチコアCPU向けの重複除去プログラム (doubri) を実装
Step 6	**品質フィルタリング**	ルールや分類器で高品質なテキストを厳選する
Step 7	**フッターの除去**	テキスト抽出では除去しきれなかったフッターを取り除く

クリーニング済み: 69,243,424ページ, 186,828,425,581文字

N Okazaki, K Hattori, H Shota, H Iida, M Ohi, K Fujii, T Nakamura, M Loem, R Yokota, S Mizuki. 2024. Building a Large Japanese Web Corpus for Large Language Models. *Conference on Language Modeling (CoLM)*.

Swallow CodeとSwallow Math (Fujii+ 2026)

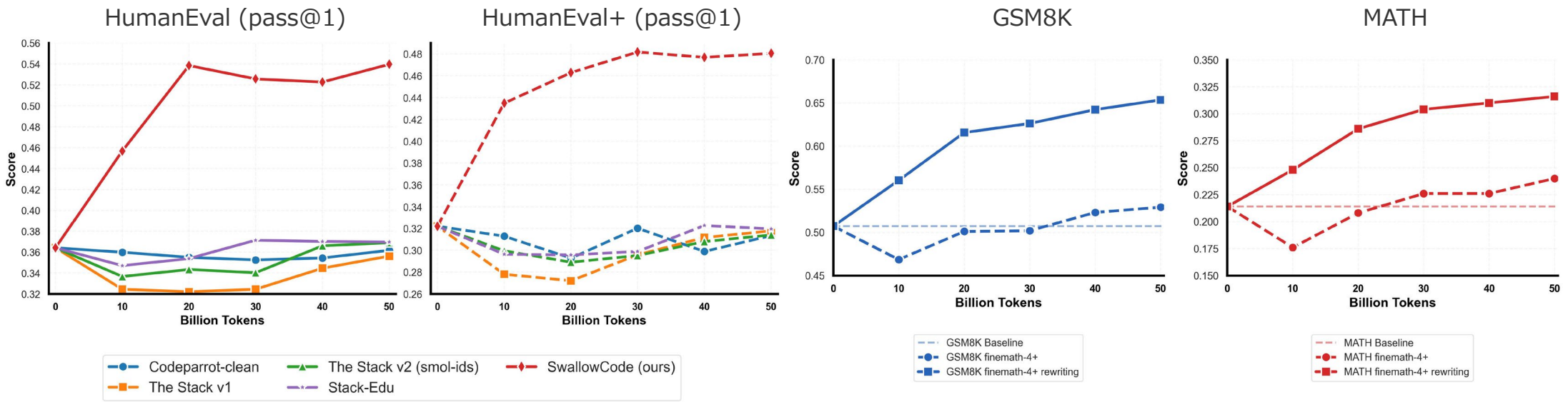
Swallow Code

データ	Stack v2 (train-smol; 36.1Bトークン)
Step1	文法エラーフィルタ (30.5Bトークン)
Step2	静的解析フィルタ (20.2Bトークン)
Step3	スタイルに沿うようにコードを書き換え
Step4	自己完結になるようにコードを書き換え

Swallow Math

データ	FineMath-4+ (9.6Bトークン)
Step1	ノイズの除去
Step2	暗黙的な情報の保管
Step3	説明の書き換え (2.3Bトークン)

(赤い箇所はLlama 3.3 70B Instructによる自動生成)



Llama 3.1 8Bを継続事前学習したときのコーディングと数学の性能

K Fujii, Y Tajima, S Mizuki, H Shimada, T Shiotani, K Saito, M Ohi, M Kawamura, T Nakamura, T Okamoto, S Ishida, K Hattori, Y Ma, H Takamura, R Yokota, J Sakuma, N Okazaki. 2026. Rewriting Pre-Training Data Boosts LLM Performance in Math and Code. *International Conference on Learning Representations (ICLR)*.

推論過程付きの指示・応答ペア

データの構築方法

- LMSYS-Chat-1Mから汎用対話の指示文、Nemotron-Post-Training-Dataset-v1からSTEM（数学・科学・コード）の設問を収集・邦訳
- gpt-oss-120b (medium) を用いて、推論過程付き応答を合成
- 問題の定式化、解法の探索、別解や前提の再検討、回答の検算を含む推論過程が得られる

ドメイン	言語	推論言語	長さ
STEM	英語	英語	2.3k
STEM	日本語	英語	2.1k
対話	英語	英語	0.8k
対話	日本語	英語	1.4k

gpt-oss-120b (medium) による合成例

We need to find minimal toral number of pens such that we can have 49 packages, each with distinct positive integer number of pens, none empty. [...] That's just the sum of the first 49 positive integers: 1+2+...+49=49*50/2 = 1225. But maybe they ask [...] So answer 1225. Check if any other constraints like each package must have at least one pen, distinct counts. So yes. Thus answer: 1225 pens.

Youmi Ma, Sakae Mizuki, Kazuki Fujii, Taishi Nakamura, Masanari Ohi, Hinari Shimada, Taihei Shiotani, Koshiro Saito, Koki Maeda, Kakeru Hattori, Takumi Okamoto, Shigeki Ishida, Rio Yokota, Hiroya Takamura, and Naoaki Okazaki. 2025. Building Instruction-Tuning Datasets from Human-Written Instructions with Open-Weight Large Language Models. In Proceedings of the Second Conference on Language Modeling (COLM), October 2025.

教師ありファインチューニング (SFT)

- ゴール
  - 対話能力の改善
- 学習データ
  - 継続事前学習と同じLMSYS-Chat-1MとNemotron-Post-Training-Dataset-v1に推論過程付きの応答を生成したもの)
  - 教師ありファインチューニングによる性能改善は限定的（継続事前学習でも同じ訓練データを入れているため)
- サイズ
  - 210M事例 (Qwen3 8B)
  - 110M事例 (その他のモデル)

GPT-OSS Swallow SFT

教師ありファインチューニング (SFT)

GPT-OSS-LMSYS-Chat-1M-Synth,
Swallow-Nemotron-Post-Training-Dataset-v1

検証可能な報酬に基づく強化学習 (RLVR)

- 学習アルゴリズムと設定
  - Group Relative Policy Optimization (GRPO) をベースに、Clip-Higher、Dynamic Sampling (Yu+ 2025)、Truncated Importance Sampling (TIS) (Yao+ 2025) などの学習安定化方策を導入
  - 推論時と学習時の MoE 専門家割り当てを一致させるRouting Replay (Zheng+ 2025) が効果的だった
  - Dolci-Think-RL-7B[1]の数学のサブセットのみを使用（ライセンスの問題がないものを厳選）
  - 回答の正誤のみで報酬を与えた
- 結果
  - 推論過程が延びるとともに、数学 (AIME) の正解率が顕著に改善した
  - 明示的に学習していないにも関わらず、コーディング (LCB) の性能も向上した
  - 継続事前学習に用いたSTEMデータによって、汎化の下地が形成された可能性があると解釈

GPT-OSS Swallow RL

検証可能な報酬による強化学習 (RLVR)

Dolci-Think-RL-7B 数学サブセット

Qiying Yu, *et al*. 2025. DAPO: An Open-Source LLM Reinforcement Learning System at Scale. NeurIPS.
Feng Yao, *et al*, 2025. Your Efficient RL Framework Secretly Brings You Off-Policy RL Training. Feng Yao's Notion.
Chujie Zheng, *et al*. 2025. Group Sequence Policy Optimization. arXiv:2507.18071
[1] https://huggingface.co/datasets/allenai/Dolci-Think-RL-7B

評価ベンチマーク (swallow-evaluation-instruct)

数学や科学、コーディングなどの高難易度ベンチマークでLLMの能力を評価

タスク	日本語	英語	評価尺度
汎用対話	日本語MT-Bench	英語MT-Bench	LLM-as-a-judge
指示追従	M-IFEval-Ja	–	正解率
日本固有の知識	JamC-QA	–	正解率（択一）
翻訳（言語生成）	WMT20 En-Ja, Ja-En	–	BLEU
常識推論	–	HellaSwag	正解率（択一）
一般教養	MMLU-ProX-Ja	MMLU-Pro	正解率（択一）
科学	★GPQA-Ja	★GPQA	正解率（択一）
数学	MATH-100-ja	★AIME 2024-25, ★MATH-500	正解率
コーディング	JHumanEval	★LiveCodeBench	Pass@1

★印は深い推論が必要なベンチマークを示す

評価設定：zero-shot推論、max_new_tokens = 32k、貪欲デコーディング（MT-Benchを除く）

https://github.com/swallow-llm/swallow-evaluation-instruct

Swallow LLM Leaderboard v2

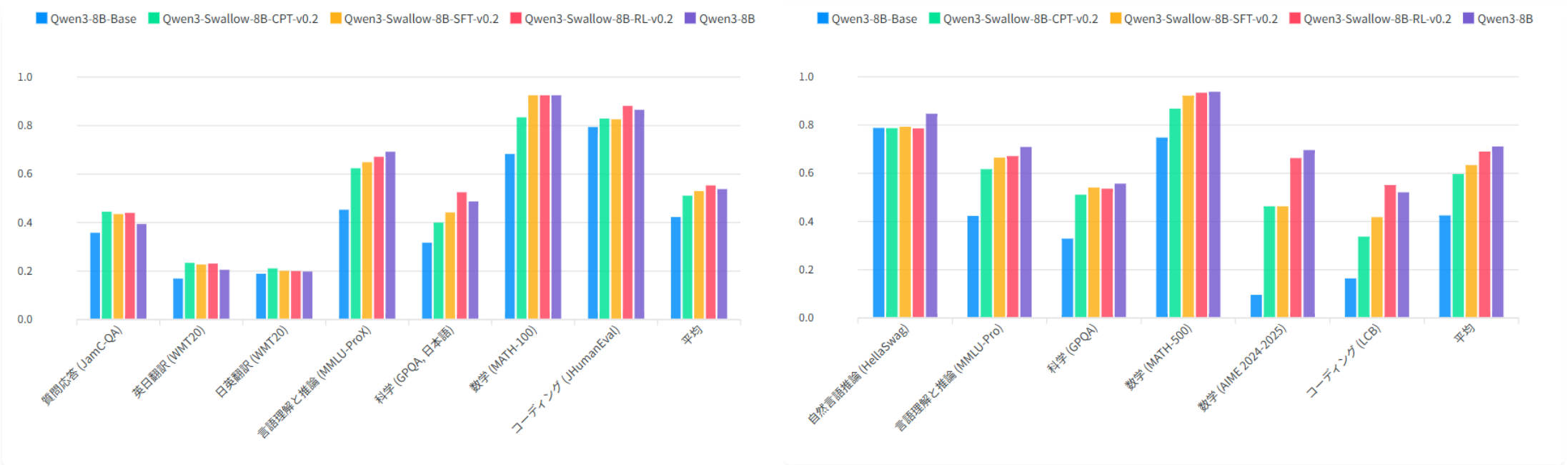
日本語に対応した様々なLLMを評価し、その結果をリーダーボードとして公開

https://swallow-llm.github.io/leaderboard/

Qwen3 Swallow 8B RLの性能

オープンなLLM（8B以下）の中で**日本語タスクで最高性能**

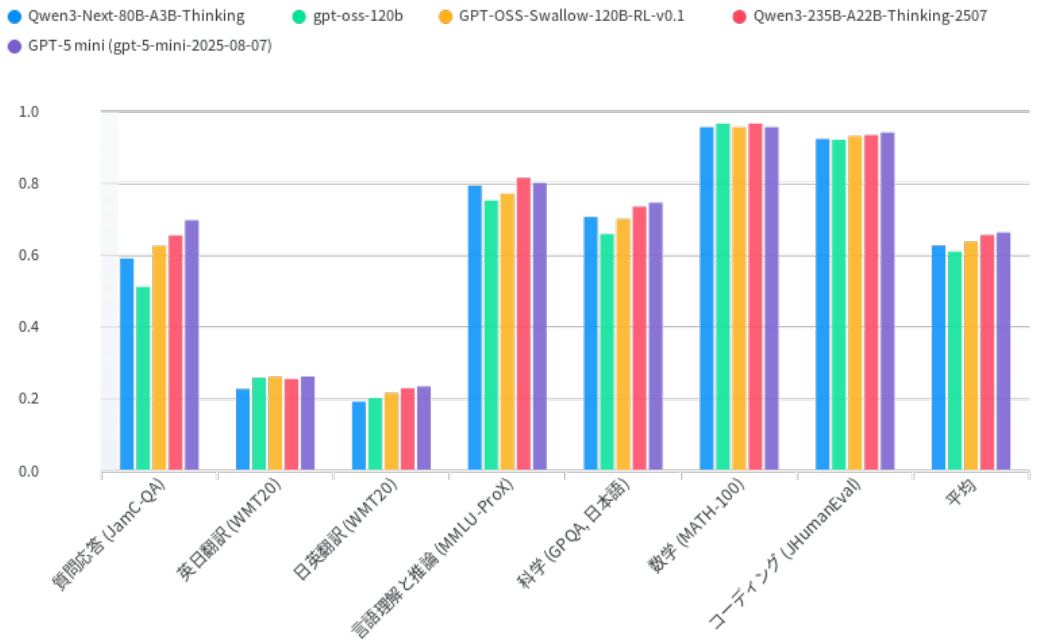
- CPT→SFT→RLで数学 (AIME) や科学 (GPQA)、コーディング (LCB) の性能が向上していくことを確認
- 日本固有の知識 (JamC-QA) や翻訳 (WMT20) の性能を引き上げることに成功（継続事前学習でモデルを構築することの目標の一つ）
- 我々の事後学習のレシピ (CFT → RL) は製造元のレシピ (Qwen3 Base → Qwen3) に匹敵する

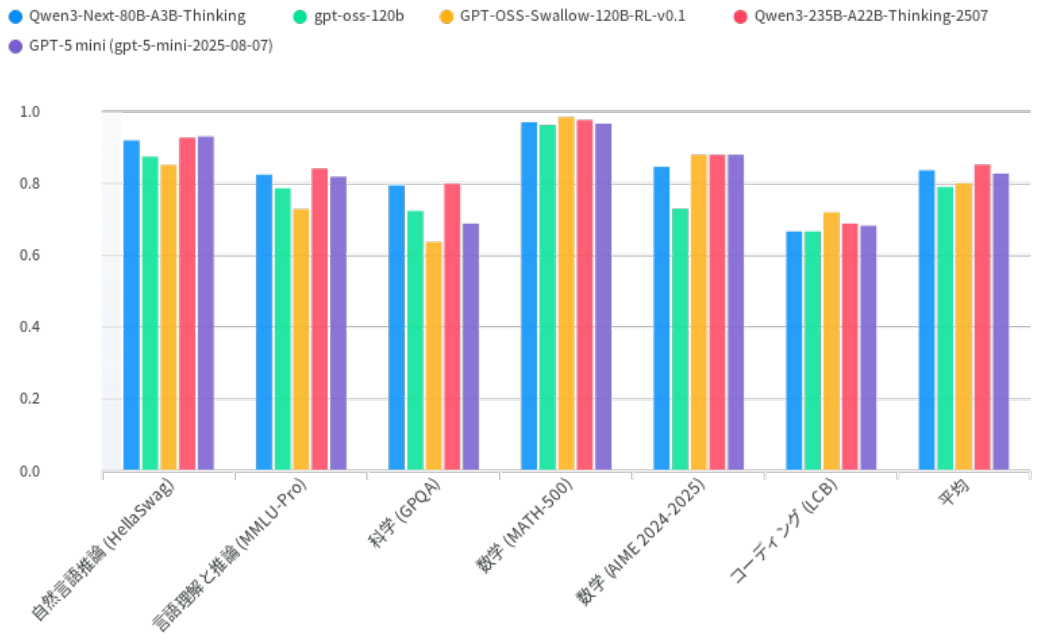


GPT-OSS Swallow 120B RLの性能

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**

- 日本固有の知識を問うJamC-QAではgpt-ossから+11ptの顕著な改善
- 数学、コード生成でもgpt-ossを上回った
- 日本語MT-Benchは0.92で最高水準
  - GPT-5.1 Thinkingを超える水準に到達し、MT-Benchでモデルの優劣判断は困難

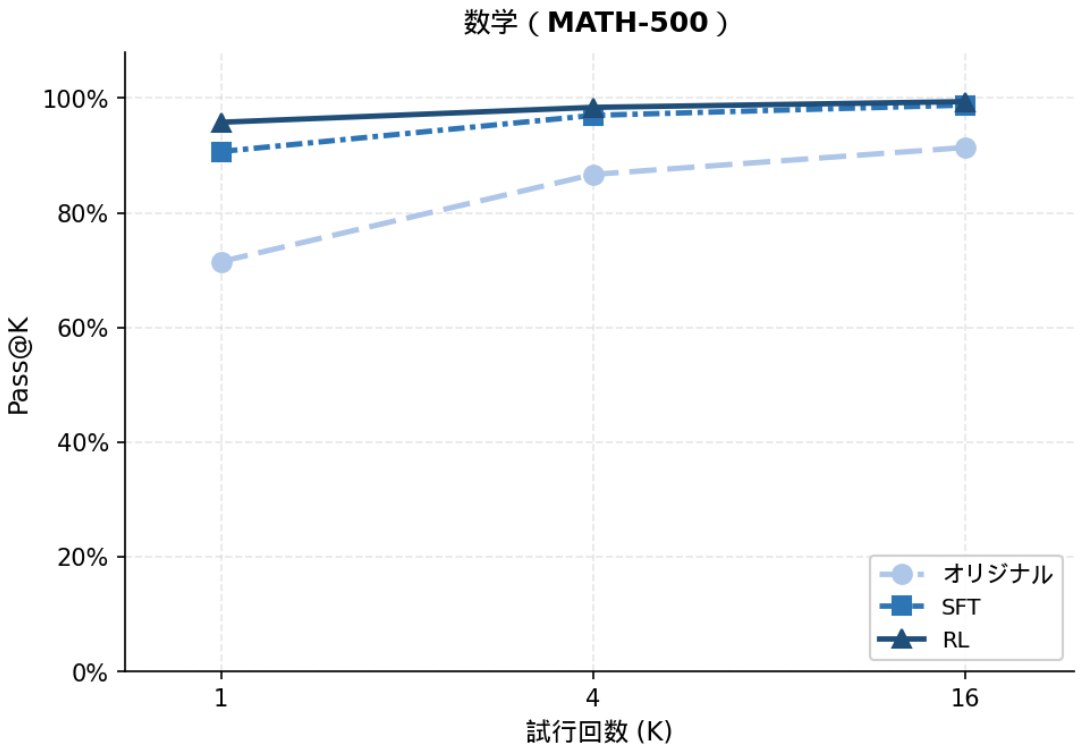


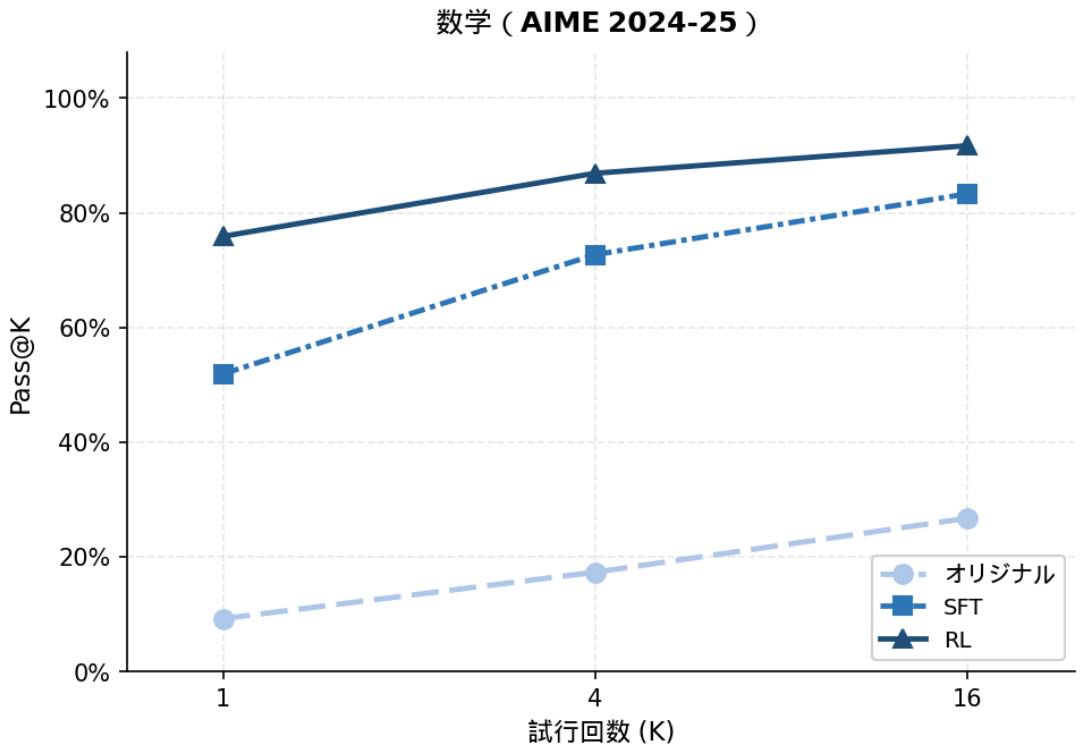


推論時スケーリングの分析（Qwen3 Swallow 8B）

RLは正解を出力しやすくするが、解けない問題を解けるようにする魔法ではない

- 問題に対して応答をk回生成させ、その中で一つでも正答があったら正解と見なし、正解率Pass@kを計算
- SFTモデルとRLモデルを比較すると、RLはPass@1でSFTに差を付けているが、kを大きくするにつれて正解率の差が減少していく





深い推論の言語間転移

① 日本語データを使わず、英語の推論過程付きデータだけで日本語STEMの性能が向上する

- 深い推論は言語間転移が期待できる

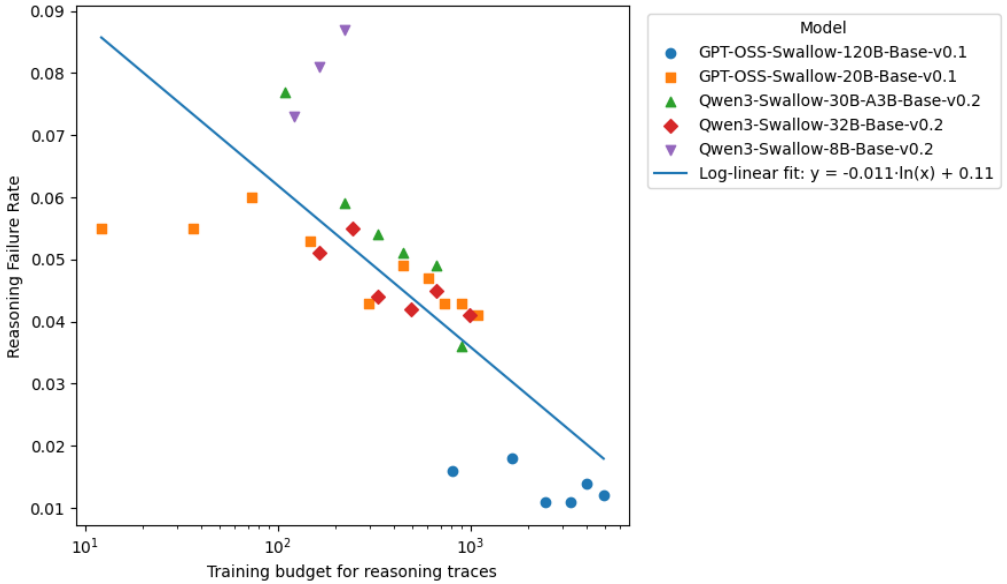
② 推論過程を除去した日本語データで学習すると、日本語で深い推論が発現しない

- 日本語で問いかけられたら深い推論を「しない」と学習してしまう
- （既存の）非推論型モデル向けの日本語のSFTデータセットを併用する場合は要注意

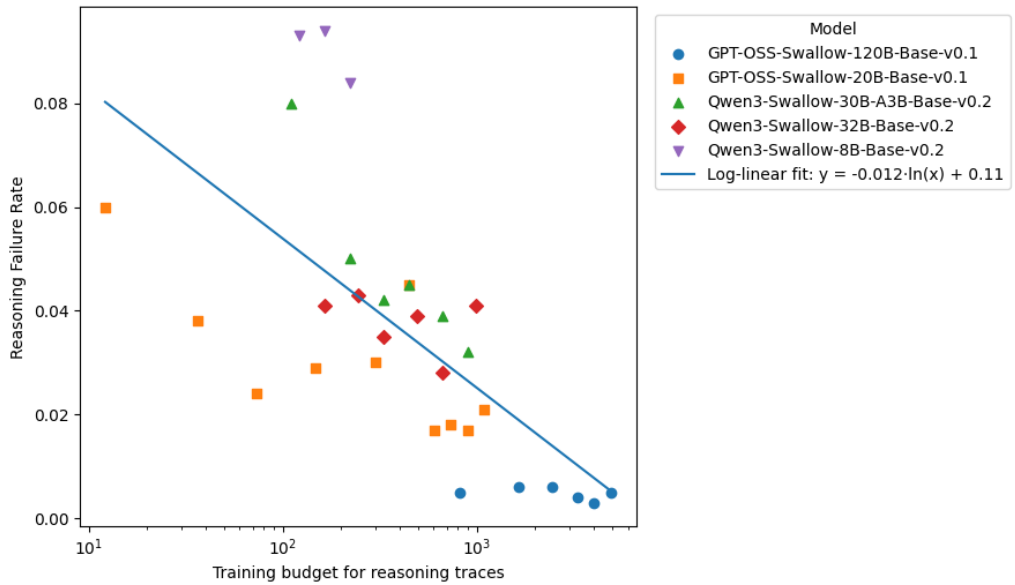
SFTの実験設定	日本語タスク			英語タスク		
	GPQA	MATH-100	JHuman-Eval	GPQA	AIME	LCB
Baseモデル (???)	0.375	0.758	0.707	0.404	0.150	0.013
① 英語STEM	**0.426**	**0.859**	**0.801**	0.472	0.379	0.296
② 英語STEM+推論過程なし日本語対話	**0.373**	**0.686**	**0.693**	0.499	0.525	0.377

推論の安定性に関するスケーリング

- CPT時に推論過程付き指示応答データを入れず、SFT時に推論過程付きデータをいきなり入れると、同じ思考を無限反復して応答に至らない「推論失敗」が頻発した（AIMEで10～40%の失敗率）
- CPT時に推論過程付き指示応答データを混ぜて学習しておくと、スケール則に沿って推論失敗率が低下し、SFT後も推論の安定が継続される
  - 性能とのトレードオフが生じやすい強化学習で推論を安定化させるよりも扱いやすい



日本語タスクの推論失敗率



英語タスクの推論失敗率

まとめ

- 東京科学大学の岡崎研究室・横田研究室、産業技術総合研究所でSwallow LLMを開発
  - Qwen3やgpt-ossをベースに、継続事前学習（CPT)、教師ありファインチューニング（SFT)、強化学習 (RL) の3段階で推論型モデルを構築
  - GPT-OSS Swallow 120Bは120B以下のオープンなLLMの中で日英タスクの両方で最高性能
- 深い推論の発現・改善・有効性に関する知見を獲得
  - RLは正解を出力しやすくするが、解けない問題を解けるようにする魔法ではない
  - 深い推論は言語間転移が期待できるが、非推論型の学習データを混ぜる場合は要注意
  - 推論過程付きデータをCPTの学習データに混ぜておくと、推論が安定化した
- 謝辞
  - 産総研政策予算プロジェクト「フィジカル領域の生成AI基盤モデルに関する研究開発」
  - 文部科学省科研費基盤A「教育的価値の高い日本語コーパスの構築による小規模言語モデル」
  - JST K-Program「大規模言語モデルのミスアライメントに対するレッドチーミング基盤」
  - 文部科学省の補助事業「生成AIモデルの透明性・信頼性の確保に向けた研究開発拠点形成」
  - LLM-jp (LLM勉強会) および大規模言語モデル研究開発センター (LLMC)