



報道関係各位

2025 年 5 月 16 日

高エネルギー加速器研究機構 (KEK)

KEK 量子場計測システム国際拠点 (WPI-QUP)

KEK 素粒子原子核研究所 (IPNS)

宇宙物理・宇宙論に特化した大規模言語モデルで、GPT-4o レベルの性能を達成

本研究成果のストーリー

● Question

高度に専門化された天文学の AI アシスタントを、大規模言語モデル (LLM) を使い、わずか 80 億のパラメータで構築することで、GPT-4o のような遙かに膨大なパラメータを持つ、計算コストの高い汎用 AI の性能に匹敵するものができるのでしょうか？

● Findings

KEK の WPI-QUP/IPNS のタイメン・デ・ハーン (Tijmen de Haan) 助教が中心となり、天文学、宇宙物理学、宇宙論、天体装置に関する公開されている約 250,000 の論文で学習させた大規模言語モデル「AstroSage-8B」を開発しました。OpenAI の GPT-4o と比べるとコストは 1000 分の一程度であるのに対し、ベンチマーク試験においてこのモデルの正答率は 80.9% に達し、OpenAI の GPT-4o の正答率をわずかに上回りました。

● Meaning

今回の研究は、特定のテーマに関して注意深く厳選されたデータを与えれば、継続的な事前学習、教師ありの追加学習と、モデル統合などの手法を組み合わせることで、小規模で、オープンウェイトなモデルが実現でき、非常に大規模なモデルを凌駕できることを示しました。これは、予算が限られている学術機関が強力な AI を導入する障壁を下げるだけでなく、自律的な研究ツールの開発にも道を開くものです。



宇宙関連の公開されている 250,000 の論文で学習させた大規模言語モデルを開発し、OpenAI の GPT-4o レベルの性能を出しました。学習方法を工夫することで低コストで高性能のツールとなりました。

概要

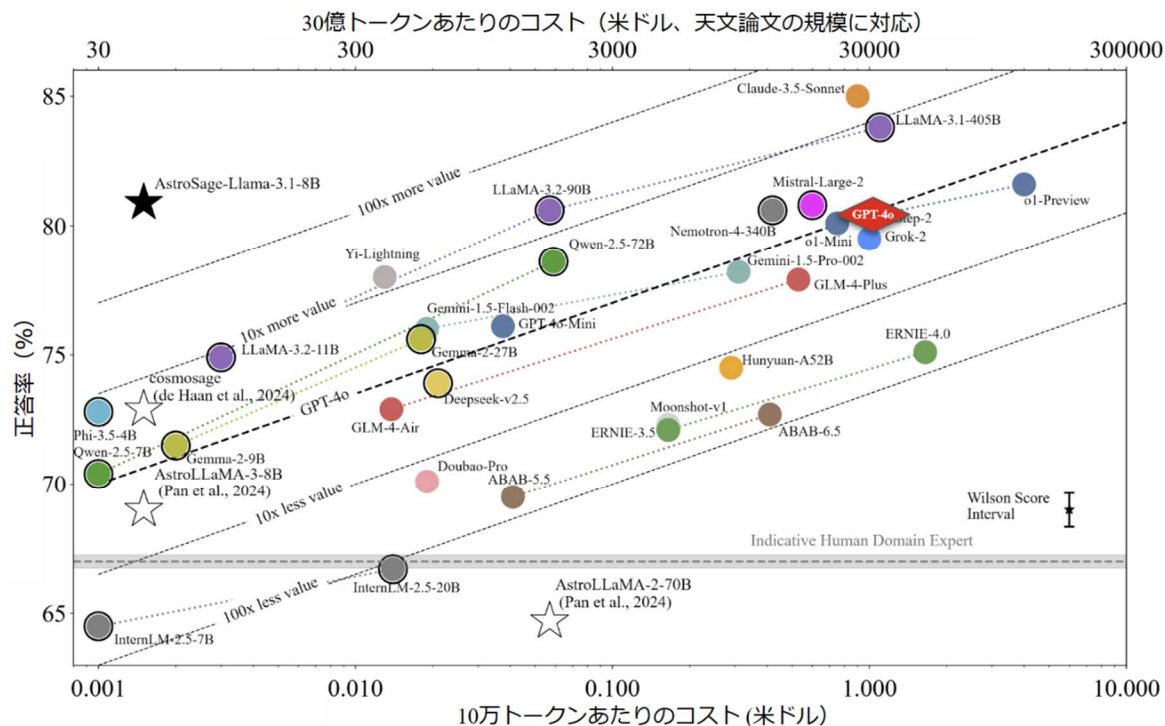
小型で高度に専門化された AI は、巨大で汎用的な AI の性能に匹敵するか、それを上回ることができるのでしょうか？これが、KEK の量子場計測システム国際拠点 (WPI-QUP) 及び素粒子原

子核研究所（IPNS）の研究者であるタイメン・デ・ハーン（Tijmen de Haan）助教と共同研究者たちが問いかけた質問です。AstroSage-8B は、この問いに「イエス」で答えます。新聞の見出しを飾るほとんどの AI は、数千億から数兆の数値「ウェイト」を含み、訓練と運用に莫大な費用をかけています。デ・ハーン博士は GPT-4o の 100 分の 1 以下である 80 億パラメータのモデルに、天文学、宇宙物理、宇宙論、宇宙科学、天体装置など、宇宙に関する領域を理解し推論することを学習させたのです。特別に訓練した AI の強力さを示す今回の結果は、2025 年 4 月 21 日に『サイエンティフィック・リポート』誌に掲載されました。

AstroSage-8B を作る工程は原理的には簡単ですが、実践的には大変なものでした。デ・ハーン博士が開発した手法[1]に沿った方法で、2007 年以降に発表された天文学と宇宙論のほとんどすべての論文（約 25 万件）を集め、それらをコンピューターで読めるような形式に変換しました。デ・ハーン博士は、エクサスケール・スーパーコンピューター「フロンティア」でモデルを学習させ、宇宙関連の知識を植え付けました。そして、このモデルにチャットボットとしての動作を教え込みました。つまり、ユーザーからの問い合わせに対して、正しく、親切で明確な回答をするようにしました。1990 年代のルールベースの AI システムとは異なり、この学習は質問例を使って進めました。何百万もの合成質問と回答のペアが LLM によって生成され、別の LLM を使ってその回答の精度をチェックしました。最高品質と判断された 880 万組の質問と回答（Q&A）例を、天文学に特化したモデルの微調整に使いました。最後に、この学習で得られたウェイトを汎用言語モデルで平均化し、Q&A の例ではカバーされていない機能、例えば複数往復の会話や、宇宙関連の領域外の質問に答える機能などを付与しました。

この 3 段階の専門的なトレーニング方法は、非常に効果的であると証明できました。4,425 問の専門家ベンチマーク[2]において、AstroSage-8B は OpenAI の GPT-4o よりもわずかに高い 80.9%の正答率で回答し、同時に約 1,000 分の 1 のコストで動作しました。モデルのウェイトをオープンライセンスで公開しているため、天文台、研究機関はもとより、大学や高校の教室などで、このトップクラスの天文学アシスタントを GPU1 基で導入することができます。研究者は、望遠鏡での観測提案書の下書き、データ解析コードのデバッグ、知識の穴埋め、新しいアイデアの検討などに利用することができ、学生は、赤方偏移、系外惑星の大気、宇宙マイクロ波背景放射などの説明を、クイズ形式でわかりやすく得ることができます。

今回開発した AstroSage-8B によって、明確に定義された科学分野にフォーカスした場合、コンパクトで手頃な価格の AI でも最大級の商用システムに匹敵する性能をだせるということを、実際に示すことができました。



ベンチマーク性能図：今回開発した AstroSage-8B(★)と他の AI モデルとのベンチマークテスト[2]での比較。それぞれのモデルでは、ほぼ斜めに引いた線に沿っており、性能があがればコストが高くなる。AstroSage の性能は GPT-4o(◆)をわずかに上回る一方、約 1000 分の 1 のコストで動作しており、目的に特化したモデルが費用対効果の高い AI を生み出すことを示している。人間の専門家の正答率は灰色の横線で示している。

研究グループ

このプロジェクトはデ・ハーン博士が率い、AstroMLab として知られる国際的なチームとの協力によります。世界中から集まったこの多様な研究者グループは、天文学者、天体物理学者、宇宙学者、自然言語処理の専門家で構成されています。

将来の展望

今回の成果はさらなる希望をもたらします。AstroSage-8B は 80 億（8B）のパラメータで優れたパフォーマンスを出せたわけで、今回の手法をより大きなモデルに拡張するとどうなるでしょうか？ 約 10 倍にパラメータを増やせば、この特化した領域で最強の LLM となるのでしょうか？「答えはもうすぐわかります」とデ・ハーン博士は言っています。「私たちは次世代モデル AstroSage-70B を現在トレーニング中で、今のところよい感じで進んでいます」。

論文情報

題名: AstroMLab 3: Achieving GPT-4o Level Performance in Astronomy with a Specialized 8B-Parameter Large Language Model

著者: Tijmen de Haan, Yuan-Sen Ting, Tirthankar Ghosal, Tuan Dung Nguyen, Alberto Accomazzi, Azton Wells, Nesar Ramachandra, Rui Pan, Zechang Sun

掲載雑誌： Journal: Scientific Reports (Nature Portfolio) **15**, 13751 (2025).

<https://www.nature.com/articles/s41598-025-97131-y>

DOI: 10.1038/s41598-025-97131-y

参考文献

[1] T. de Haan, “Cosmosage: A Natural-Language Assistant for Cosmology,” *Astronomy & Computing* 51 (2025) 100934.

[2] Y.-S. Ting, T. de Haan, et al., “AstroMLab 1: Who Wins Astronomy Jeopardy!?,” *Astronomy & Computing* 51 (2025) 100893.

お問い合わせ先

<研究内容に関すること>

大学共同利用機関法人 高エネルギー加速器研究機構

量子場計測システム国際拠点 (WPI-QUP)

e-mail: qup_pr@ml.post.kek.jp

素粒子原子核研究所 (IPNS)

Tel: 029-864-5408

e-mail: ipns-pr@ml.post.kek.jp

<報道担当>

大学共同利用機関法人 高エネルギー加速器研究機構 広報室

Tel: 029-879-6047

e-mail: press@kek.jp