

(1148-2)

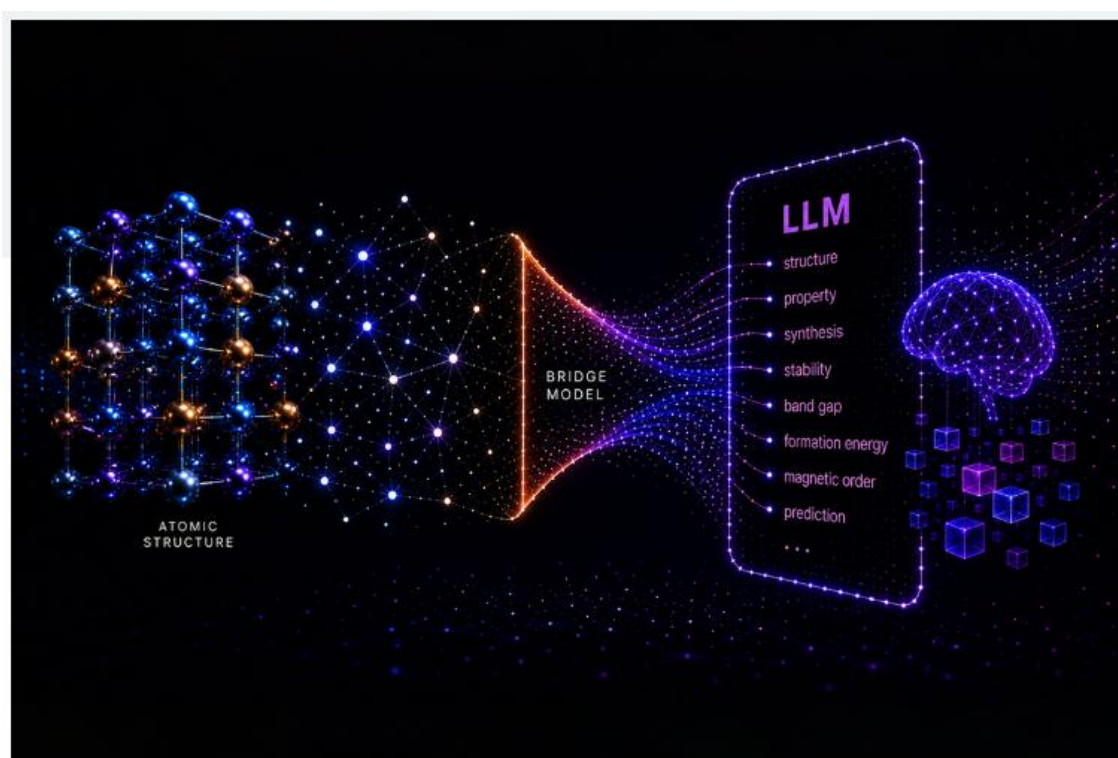
【ロボット・AI 技術分野】

仮訳

AI による科学言語の「理解」を支援する新しい MatterChat モデル (米国)

By Linda Vu

2026 年 5 月 18 日



MatterChat は、原子間の複雑な 3D の力をテキストベースの AI が理解できるように支援する特殊な架け橋として機能し、新しい材料の発見のための強力なツールへと AI を進化させる。

(Credit: Yingheng Tang/Zhi Jackie Yao/Berkeley Lab)

ローレンスバークレー研究所が、特殊な物理モデルと対話的な大規模言語モデル(LLMs)を橋渡しする、汎用的な AI「ブリッジモデル」の MatterChat を開発した。これは、複雑な科学データを大規模言語モデル(LLMs)に統合するパワフルな新しい青写真を提供するもので、原子構造の言語の理解を通じた材料発見の加速に使用される。

メール作成からコンピューターコードの生成まで、日常生活に広く普及している人工知能(AI)の多くは、テキストという一つの領域を極めることに成功している。しかし、結晶中の原子の複雑な格子構造といった物理世界の高解像度三次元データに依存する物理学分野は大きな盲点となっている。科学における AI 活用の可能性を実現するには、こうしたデータ駆動型のテキストモデルが、物理ベースのモデルとシームレスに「対話」できるようにする必要がある。

ローレンスバークレー国立研究所（バークレー研究所）が開発した新しい AI フレームワークの MatterChat は、特殊な「ブリッジ」を作ることでこの問題を解決する。MatterChat は、大規模言語モデル(LLMs)の対話能力と、原子間の複雑な物理力である「原子間ポテンシャル」をモデル化する物理ベースの AI を結びつける。MatterChat はすでに、材料特性の予測において GPT-4 のような汎用 AI ツールの能力を大幅に超えており、これが確かな知見を提供し、新規材料の合成手順を段階的に生成する強力な研究パートナーとして機能することで、科学的発見を加速できると研究チームは期待している。

この研究成果をまとめた論文は、[*Nature Machine Intelligence*](#) 誌に掲載されている。

「従来のシミュレーションは、材料科学に必要な物理的厳密性を提供できるものの、計算コストが高すぎてハイスループットスクリーニングには適していません。一方、LLMs は迅速な知識合成に優れてはいるものの、原子座標から直接材料を読み取る『構造的なビジョン』が本質的に欠けています」と、バークレー研究所応用数学・計算研究部門(AMCR)の博士研究員であり、この論文の筆頭著者である Yingheng Tang 氏は言う。「MatterChat は、このジレンマを解決するために開発されました。LLMs に構造的な『ビジョン』を与えることで、複雑な現実世界の材料課題を解決するための、LLMs の潜在能力を最大限に引き出すことができるのです」。

言語モデルを活用して材料科学における複雑な課題を解決

MatterChat の開発にあたり、バークレー研究所のチームは、視覚的質問応答(VQA)やテキストから画像への変換(T2I)といった技術から着想を得た。これらのタスクでは、AI は高度なテキスト概念を視覚的な画像に変換したり、その逆を行ったりする必要がある。これには、根本的に異なる 2 つのデータ形式を「橋渡し」するツールの開発が必要となる。

研究者らは、このコンセプトを物理科学に応用した。MatterChat により、LLMs の一般的な知識と、科学的な原子間ポテンシャルに含まれる原子スケールの世界に関する深い理解を効果的に結びつける「橋渡しモデル」を構築した。

これまで、材料科学の問題解決に LLMs を用いる研究者は、生のデータファイルを単なるテキスト文字列として入力する必要があった。これは、部品リストのみを基に複雑な 3D エンジンを AI に理解させようとするようなものである。LLM は部品名を読み取ることはできるが、原子が空間内でどのように配置されているかを「見る」ことができない。MatterChat は、数百万もの結晶構造で事前学習された特殊な AI ブリッジモデルと LLM を訓練し、LLM による現実的世界の表現と原子間ポテンシャルによるそれとを整合させてこの問題を解決する。

このブリッジモデルは、物理的な知見を LLM が実際に理解できる形式に変換することができる。LLM にこうした「科学的な視点」、つまり AI 用語でいうところの科学的な「帰納的バイアス」を付与することで、バークレー研究所のチームは、熱安定性の予測や電子バンドギャップの解析といった複雑な材料課題に対して、根拠に基づいた科学的知見を提供できる強力な研究ツールへと LLM を変貌させた。

「原子は物理的な空間に存在するものと私たちは考えますが、機械学習の観点から見ると、原子は高次元ユークリッド空間における非常に複雑な構造の多様体に存在するベクトルに過ぎません。そしてもちろん、原子に関する私たちの考えを表現するために用いる文章や段落についても同じことが言えます」と、論文の共著者でありバークレー研究所 Scientific Data Division (SDD) AI Initiative Research Lead の Michael Mahoney 氏は説明する。「このブリッジモデルは、これらの 2 つの構造が互いに『対話』できるようにするものです」。

この一般的なアプローチの概念実証として、研究チームは [Materials Project](#) から収集した約 14 万 3000 個の安定した原子構造と、それに対応する物理的特性を一組にしたデータセットを用いてこのブリッジモデルを訓練した。この訓練データは Materials Project の API を用いて自動的に構築され、マイクロエレクトロニクス設計に不可欠な形成エネルギーやバンドギャップ等の特性で意図的に強化されており、MatterChat が材料の原子構造と機能的性能を結びつける複雑なパターンを学習できるようにする。

汎用的な LLMs から他の専門的な科学 AI 手法に及ぶ、様々な AI システムとの性能評価で MatterChat を検証した結果、MatterChat は幅広いタスクにおいて競合システムを超える性能を示した。このブリッジモデルは、材料の種類分類においてより高い精

度を発揮し、数値特性の予測精度も優れていた。例えば、高容量エネルギー貯蔵から次世代コンピューターチップに至るまで、新しいエレクトロニクス設計に不可欠な特性である材料のバンドギャップの予測において、**MatterChat**は卓越した性能を示した。

「大規模な AI モデルをゼロから構築する必要がないため、私たちの設計は非常に効率的です」と、論文共著者であり、バークレー研究所 AMCR の研究員である Zhi (Jackie) Yao 氏は言う。「私たちは、材料物理学向けの構造エンコーダーとオープンソースの LLM の、パワフルで事前学習済みの 2 つのモデルをそのまま利用しています。実際に学習させているのは、これらのモデルをつなぐこの軽量の『ブリッジモデル』のみです。これは、自動車工場全体を建設することと、世界最高水準のエンジンと世界最高水準のナビゲーションシステムとを接続するスマートアダプターを設計することとの違いに相当します。このアプローチは計算効率が高いだけでなく、システムをモジュール化するため、将来的にコンポーネントを容易にアップグレードしたり、ブリッジモデルを他の科学分野に適応させたりすることができます」。

重要な点は、このモジュール設計が、バークレー研究所や米国エネルギー省(DOE)のような機関が、急成長する AI 分野においていかにして非常に価値の高いニッチ市場を切り開いているかを明示していることである。シリコンバレーの巨大テクノロジー企業と競合してより大規模な言語モデルをゼロから構築するのではなく、商用 AI をハードコア・サイエンスに役立てるための特殊な橋渡しに焦点を当てている。

MatterChatの基盤となるブリッジモデルのアプローチは将来的な互換性を備えているため、これらの並行するイノベーションの流れを最大限に活用できる最適な位置にある。Mahoney 氏が指摘したように、「産業界は今後も優れた LLMs を開発し続け、各分野の科学者や組織は新たなデータを生成し続けると予想します。科学的機械学習の重要な点は、単に今日のデータで問題を解決することではなく、科学分野であろうと LLMs であろうと、桁違いに大量のデータに対応できる汎用的な手法を開発することです」。

Yao 氏によると、バークレー研究所の Laboratory Directed Research & Development (LDRD)プログラムの資金援助を受けて開発・改良された **MatterChat** プロジェクトは、今後その機能を拡大していく予定であるという。フェルミ国立加速器研究所との共同研究において、**MatterChat** は既に米国エネルギー省の [Genesis Mission](#) プロジェクトの Accelerating eXtreme Environment Specs-to-Silicon (AXESS)に貢献している。AXESS は、高度な 3D 集積回路(チップレット)と AI 駆動型データ分析を用いて次世代

の高速・耐放射線検出器の開発を加速し、素粒子物理実験での活用を目指している。

MatterChat の成功には、LDRD のサポートに加え、バークレー研究所の国立エネルギー研究科学計算センター(NERSC)のスーパーコンピューティングリソースも貢献している。「NERSC には大変感謝しています。[AI for Science プログラム](#)を通じて *Perlmutter* スーパーコンピューターを利用できなければ、この研究は実現しなかったでしょう」と Tang 氏は言う。当時 NERSC のポストドク研究員で本研究論文の主要共著者の Wenbin Xu 氏、また、バークレー研究所の SDD のリサーチ・サイエンティストである本研究論文共著者の Benjamin Erichson 氏も、AI for Science における AMCR-SDD の連携の利点を強調している。

訳：NEDO（担当 イノベーション戦略センター）

出典：本資料は、ローレンスバークレー国立研究所(LBNL)の記事“New MatterChat Model Helps AI to ‘See’ the Language of Science”

(<https://newscenter.lbl.gov/2026/05/18/new-matterchat-model-helps-ai-to-see-the-language-of-science/>) を翻訳したものである。