



### Neal R. Brodnik

Department of Mechanical Engineering,  
University of California Santa Barbara,  
Santa Barbara, CA 93106  
e-mail: nbrodnik@ucsb.edu

### Samuel Carton

Department of Computer Science,  
University of New Hampshire,  
Durham, NH 03824  
e-mail: samuel.carton@unh.edu

### Caelin Muir

Materials Department,  
University of California Santa Barbara,  
Santa Barbara, CA 93106  
e-mail: muir@ucsb.edu

### Satanu Ghosh

Department of Computer Science,  
University of New Hampshire,  
Durham, NH 03824  
e-mail: satanu.ghosh@unh.edu

### Doug Downey

Allen Institute for Artificial Intelligence,  
Computer Science & Engineering,  
Northwestern University,  
Evanston, IL 60208  
e-mail: d-downey@northwestern.edu

### McLean P. Echlin

Materials Department,  
University of California Santa Barbara,  
Santa Barbara, CA 93106  
e-mail: mechlin@ucsb.edu

### Tresa M. Pollock

Materials Department,  
University of California Santa Barbara,  
Santa Barbara, CA 93106  
e-mail: tresap@ucsb.edu

### Samantha Daly<sup>1</sup>

Department of Mechanical Engineering,  
University of California Santa Barbara,  
Santa Barbara, CA 93106  
e-mail: samdaly@ucsb.edu

# Perspective: Large Language Models in Applied Mechanics

*Large language models (LLMs), such as ChatGPT and PaLM, are able to perform sophisticated text comprehension and generation tasks with little or no training. Alongside their broader societal impacts, these capabilities carry great promise for the physical sciences, including applied mechanics. We present a summary of recent developments in these models, their application to mechanics and adjacent fields, and a perspective on their future use in applied mechanics, taking into account their limitations and the unique challenges of the field. [DOI: 10.1115/1.4062773]*

*Keywords: computational mechanics, mechanical properties of materials, machine learning, artificial intelligence, natural language processing*

## What Is a Large Language Model?

In natural language processing (NLP), a *language model* is a probabilistic model of sequences of words. When conditioned on

a partial text sequence, a language model learns from its training corpus to predict the most likely missing text. By learning to do this task well, such models implicitly absorb linguistic rules, norms, idioms, and even world knowledge from the text on which they are trained. The era of “large” language models (LLMs) began with the 2017/2018 publication of the transformer architecture [1] and 110M- and 345M-parameter BERT models [2] (Fig. 1).

<sup>1</sup>Corresponding author.

Manuscript received May 19, 2023; final manuscript received June 7, 2023; published online July 17, 2023. Assoc. Editor: Pradeep Sharma.

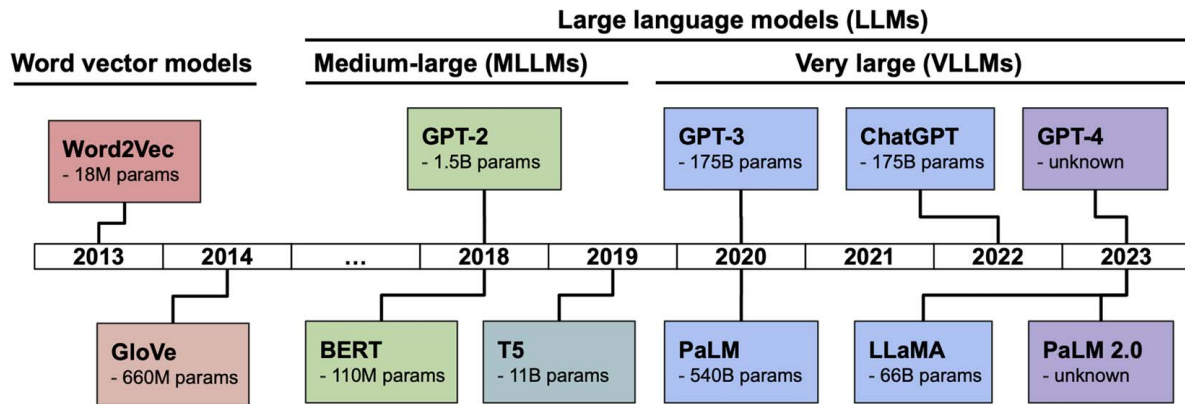


Fig. 1 Timeline of recent language models and their approximate categorization into medium and very large

When a model like BERT is pretrained on a large text corpus with an unsupervised language modeling objective, it serves as a rich starting point for further training (fine-tuning) on specific supervised learning tasks, such as information extraction or text classification; analogous to learning a language before learning to perform a specific task in that language. Until recently, the paradigm of “general-purpose language model pretraining followed by task-specific fine-tuning” has dominated NLP. However, more recently, larger LLMs have begun to obviate the need for fine-tuning altogether.

### What Are the Emergent Capabilities of Large Language Models?

Recently, *very* large language models (VLLMs) such as 540B-parameter PaLM [3] and the 175B+ GPT 3/4 series [4,5] have demonstrated the ability to perform *zero-* and *few-shot* learning (Fig. 2) on the basis of pretraining alone. In consumer-facing models such as ChatGPT, this consists of both text-completion-based language model pretraining and additional, yet still general-purpose, human-in-the-loop optimization for instruction-taking [6]. During zero-shot learning (applied to Ref. [7] in Fig. 2), the model is prompted with the textual description of a task, and it generates answer text based solely on the linguistic and real-world knowledge it absorbed during pretraining. In few-shot learning, this prompt is accompanied by a small number of *exemplars* (shown using Refs. [7,8] in Fig. 2) which demonstrate for the model how to perform the task [4]. Increasingly, these learning paradigms outperform traditionally fine-tuned models (and humans) on a wide variety of linguistic tasks [5]. These emergent capabilities are exciting for science because they vastly lower the barrier to effective scientific data mining. They require no large datasets of expert annotations, and eliminate the time, computational power, and expertise associated with traditional model training. Instead, they at most need a small set of expert exemplars and no additional fine-tuning.

These very large models have flaws. Because they are optimized for linguistic plausibility, they have a tendency to *hallucinate*, or fabricate, fictitious information, and they can make reasoning errors when faced with tasks involving multi-step logical inference [5]. They also tend to produce “boring” output that represents a middle-of-the-road consensus of information absorbed from their training data, which limits their ability to, for example, produce novel scientific hypotheses [9]. The emerging field of *prompt engineering* [10] explores prompting strategies for mitigating these issues, such as “chain of thought prompting” [11] which explicitly demonstrates step-by-step reasoning to avoid logical errors or “retrieval-augmented generation” which incorporates retrieved documents into the input prompt [12] in order to discourage hallucination. This latter approach is an example of “augmented generation,”

where a LLM is optimized to operate in concert with additional functions such as web search, calculation, or a calendar [13,14].

At the time of writing, there is a distinction between “medium-large” language models (MLLMs) such as BERT and its variants, and “VLLMs” such as PaLM and GPT-3/4 (see Fig. 3). MLLMs generally are not capable of zero- or few-shot learning, and must be fine-tuned on training data. However, they can be purpose-pretrained on specific corpora (e.g., on a corpus of mechanics literature) and can be fine-tuned in innovative ways, such as for integration with scientific knowledge bases (e.g., Ref. [15]). VLLMs generally are capable of zero- and few-shot learning, but are too large to purpose-pretrain or easily fine-tune on commercially available compute (with the partial exception of GPT-3, whose application programming interface (API) allows for opaque, proprietary, and limited fine-tuning). In this perspective, we use the term MLLM to refer to LLMs small enough to allow (and require) traditional fine-tuning, where VLLMs are more suited for zero- or few-shot learning, with at-best limited fine-tuning ability. Another research direction in LLMs studies the interplay between large general-purpose models and smaller, more focused models, to leverage the advantages of both LLM types [16].

The line between MLLM and VLLM is ill-defined and liable to change in the next few years, as models and computational resources continue to advance. For most teams, this line lies between the 11B-parameter T5 [17], which is often fine-tuned, and the 120B-parameter Galactica [18], which is designed for inference only (no fine-tuning) on the highest-end commercially available graphical processing units (GPUs).

### How Have Large Language Models Been Applied in Mechanics Related Fields?

A collection of recent LLMs have been pretrained specifically on corpora of scientific literature and found to perform better on scientific NLP tasks than ones trained on a general-purpose text corpus. Examples include MLLMs like SciBERT [19] and ScholarBERT [20], which outperform the original BERT on such tasks, and the VLLM Galactica [18], which outperforms GPT-3. This idea of domain-specific pretraining has been applied even more narrowly to materials science. Examples include MatBERT [21], which is pretrained on a corpus of materials papers and, when fine-tuned, outperforms SciBERT on materials-specific NLP tasks. Similar models include MatSciBERT [22] and MaterialBERT [23], as well as even more focused models such as BatteryBERT [24], OpticalBERT [25], and TransPolymer [26].

Parallel work has explored how to use LLMs as “soft knowledge bases,” focusing on what factual information they absorb through their pretraining and how to extract it. This idea is key to their use in scientific applications, which are heavily dependent on factual knowledge. Petroni et al. [27] use “fill-in-the-blank”

Prompt	One-shot exemplar
<p>I am an intelligent assistant capable of understanding material mechanics. I will be reading a document created from a paper on material mechanics. I will answer questions about the paper by extracting words from the given document. I will not generate any new words. I will output in a tabular format.</p> <p>Question: Describe the goal of this study, what was tested, the results of those tests, and why those results occurred</p> <p>[Optional one-shot exemplar (shown to right)]</p> <p>The document is: [Introduction and results from Szyniszewski et al. (2020)]</p>	<p>The document is: [Introduction and results from Cresswell-Boyes et al. (2022)]</p> <p>Goal: The aim of this study is to 3D print typodont teeth with materials developed to mimic both the morphology and mechanical response of natural teeth. What was tested: X-ray microtomography (XMT) is used here to image natural teeth to accurately map the geometry of the samples, which are 3D printed. A force measuring system was developed for mechanical cutting evaluation. Results: XMT is adept at fully describing the morphology of teeth ... demonstrating a difference when altering, not only the reinforcement material but also the wt. % Why results occurred: This discrepancy in forces may explain why undergraduate students dislike ... the additional force required to cut typodont teeth is expected to be due to the mechanical performance of the materials under the cutting action rather than the clogging of the bur.</p>
Zero-shot output	One-shot output
<p>Goal of the study: Create a new metallic-ceramic, hierarchical structure that is susceptible to internal vibrations under localized loads. What was tested: Resistance of the metallic-ceramic hierarchical structure to extreme loads such as angle grinder, power drill, and water jet cutter. Results: 1. Angle grinder: The cutting discs experienced high wear and became ineffective after 60-65 seconds. 2. Power drill: The drilling progress was arrested when the drill encountered a ceramic sphere. 3. Water jet cutter: The abrasive jet was deflected from the ceramic sphere, and the cutting rate was significantly slower compared to a solid steel plate. Why those results occurred: 1. Vibrational interface mechanism: The contact with the ceramic segment produced high-frequency, out-of-plane vibrations, which interfered with the cutting tools. 2. Abrasive wear: Fragmentation of the ceramic spheres into particulate matter created a sandpaper-like, vibrating interface with the cutting disc. 3. Contrast between stiffness: The difference in stiffness between the ceramics and cellular metallic foam matrix, as well as the susceptibility of the ceramic spheres to fragmentation, were effective against the waterjet cutter.</p>	<p>Goal: The goal of this study is to create a new metallic-ceramic, hierarchical structure that is susceptible to internal vibrations under localized loads and resistant to extreme loads such as angle grinders, power drills, and water jet cutters. What was tested: The study involved producing sandwich plate specimens made of a cellular aluminum core with an orthogonal layout of ceramic spheres and steel alloy faceplates. The specimens were subjected to compressive loading, angle grinder attacks, power drill penetration, and water jet cutting. Results: The cellular structure showed significant deformability, exceeding 20% of engineering strain. The angle grinder achieved only a partial incision and experienced high wear. The vibrational mechanism was equally effective against the power drill, arresting drilling progress when encountering a ceramic sphere. The water jet cutter was deflected from the first encountered ceramic sphere and accompanied by reverse splashing toward the nozzle. Why results occurred: The resistance to extreme loads is due to the contrast between the stiffness of the ceramics and the cellular metallic foam matrix, as well as the susceptibility of the ceramic spheres to fragmentation. The vibrational interface mechanism is not sensitive to the aspect ratio of the rotating cutting tools. The abrasive wear of the cutting disc is due to the fragmentation of the ceramic spheres into particulate matter. The curvature of the ceramic obstacle in the water jet cutter test was responsible for widening the jet, which slowed its velocity by two orders of magnitude.</p>

**Fig. 2** Example of GPT-4 (highlighted output) performing zero-shot and one-shot semi-structured information extraction on Ref. [7]. In zero-shot, GPT-4 focuses on general material properties (including an incorrect engineering strain description) and struggles with causal relationships that drive non-cuttability. In one-shot, GPT-4 is provided with a manually extracted exemplar from a study on dental composite cuttability [8], which leads to a more narrative response that better captures property descriptions around the design goal. Note that both table-like and narrative output are possible, and model output will depend on the exemplar structure.

prompts (“clozures”) to demonstrate that even the original BERT model encodes meaningful factual information (e.g., the birthplace of Dante Alighieri). This approach has become a template for zero-shot prompting of larger models as well. In the evaluation of Galactica, the model achieved 43% accuracy at predicting chemical reaction products from reagents [18]. Kandpal et al. [28] use similar prompts to show that models struggle to learn facts that are not well-supported in their training data.

When considering LLMs as “soft” knowledge bases, a related line of work asks how to combine them with traditional “hard” knowledge bases such as material property datasets. Safavi et al. combined vector representations from traditional knowledge bases with those produced by language models for improved link-prediction within these knowledge bases [15]. Similar work has primarily been pursued in the biomedical domain, aligning LLM-derived representations with data from biomedical databases [29–31].

The idea that a language model can encode useful world knowledge precedes contemporary models. In materials, the use of word vector models [32,33] by Tshitoyan et al. [34] on thermoelectric compounds attracted significant attention because it showed that representations extracted from even such a simple model can predict the thermoelectric properties of compounds. This approach has been applied to exploration of the properties of polymers [35], multi-principal element alloys [36], and fuel cells [37].

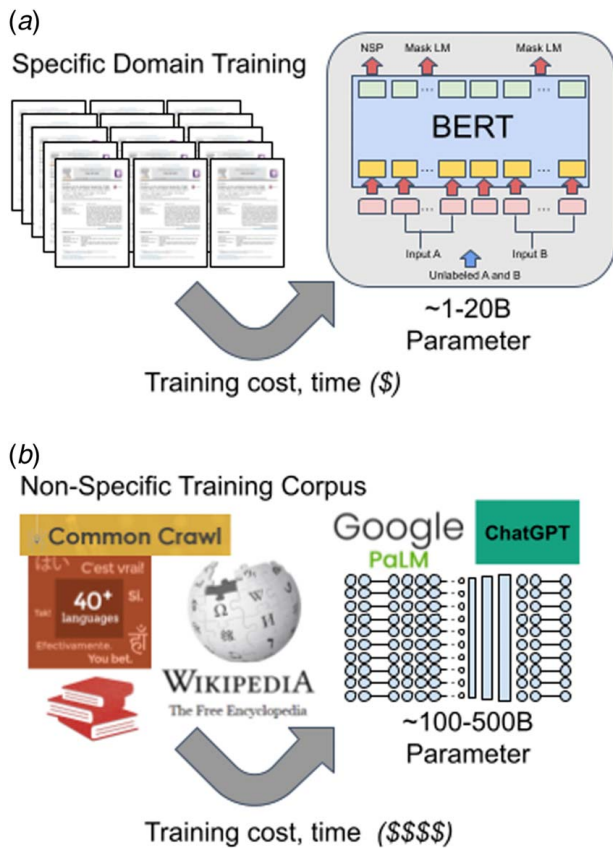
Domain-specific pretraining and fine-tuning for alignment between LLMs and structured knowledge bases lend themselves to use in MLLMs, as VLLMs require more time, computational infrastructure, and data than is available to anyone except a small handful of companies. Galactica is an exception here, as it was released by Meta for general-purpose scientific applications. However, the explosion of interest in VLLMs has also touched on their use in science, with scientists in varying fields expressing

both interest in the potential and concern for the pitfalls of general-purpose large generative models [38,39]. This excitement has spread to material mechanics, where developments such as Refs. [40–42] take advantage of OpenAI’s proprietary tuning API to demonstrate that minimal supervised fine-tuning enables VLLMs like GPT-3 to make high-fidelity property predictions. For VLLMs that cannot be tuned, Polak and Morgan [43] demonstrate a zero-shot prompting strategy for material information extraction from the published literature.

The use of LLMs in mechanics has been sparse. Independent of its use in LLMs, the Transformer architecture has attracted attention as a means for modeling of material properties [44–46]. There is also excitement about the use of LLMs for large-scale text mining in mechanics and materials science [47]. Some pre-LLM, traditional NLP approaches have been applied to mechanics problems [48–52], but to the best of our knowledge, the only instance of LLMs explicitly used for mechanics is Ref. [53], which applies a supervised BERT model to describe textual mechanics problems to other solver methods.

### What Are the Challenges of Applying Large Language Models to Mechanics?

The use of LLMs in applied mechanics is challenging due to the focus in mechanics on structural design and phenomenological description across a wide range of topics. Traditionally, MLLMs (which have dominated the recent study of NLP in science) have succeeded in fields where the knowledge represented in the literature can be efficiently extracted and organized into structured schemas (e.g., ontologies), which can then be used as training datasets. For example, in materials science, descriptors such as the stoichiometric



**Fig. 3** There are significant differences between medium- and very large language models (MLLMs and VLLMs), including that they require substantially different amounts of cost and time to train. MLLMs can more readily be focused on domain-specific literature (e.g., focused scholarly literature) and trained by research group-scaled entities, whereas VLLMs require substantial investment and infrastructure and have more generally focused training (Common Crawl databases, books, and Wikipedia) that is not domain specific. The differences in domain focused training for each model type has implications for hallucination. Current MLLMs are around 1–20 billion degrees-of-freedom (DOFs) [17] compared to VLLMs that are currently hundreds of billions of DOF [5]. (a) MLLM and (b) VLLM.

formulas of alloys can be readily extracted and linked to physical properties in the form of structured databases, such as those used for CALPHAD (e.g. Thermo-Calc, Pandat) and density functional theory (DFT) (e.g., OQMD: Open Quantum Materials Database, Materials Project). This formulaic nature that enables database construction also allows for efficient annotation of training and evaluation data for NLP models. For example, Tshityan et al. [34] validate vector representations of material formulas against a database of thermoelectric figures-of-merit.

One example of how schema creation can be difficult in mechanics is illustrated by the most downloaded Nature Materials paper of 2021, Ref. [7], which describes a non-cuttability composite composed of ceramic spheres embedded in an aluminum foam matrix capped with steel face plates. If one wanted to create a large-scale database linking composites to their properties, it would be challenging to design a schema flexible enough to represent all possible composite structures and connect those structures to emergent mechanical phenomena, such as the strain rate and local resonance effects that drive the non-cuttability seen in Ref. [7].

Beyond characterizing materials, many papers in applied mechanics are standalone investigations of context-dependent phenomena. An example is Ref. [54], which is a systematic study of elastically bistable ring origami, and one of the top five “most read” papers in the *Journal of Applied Mechanics* in May 2023.

Elastic bistability is a useful phenomenon, but it occurs in many contexts beyond ring origami, such as clamped beams [55] and metamaterials [56]. Similar to representing composite behavior, it would be challenging to design a database schema capable of unifying the observed behaviors and these three disparate examinations of elastic bistability.

The diverse, difficult-to-categorize nature of applied mechanics research represents a key challenge for traditional NLP methods, including MLLMs. However, by reducing the need for traditional fine-tuning (and the large-scale databases it requires), VLLMs represent a possible and powerful means of overcoming these challenges.

## How Can Large Language Models Be Applied in Mechanics?

We discuss several key emergent capabilities of LLMs that represent promising directions for their use in applied mechanics.

**Zero- and Few-Shot Information Extraction.** The first and perhaps most immediate such capability is information extraction. With little or no supervision, VLLMs can perform sophisticated reading comprehension tasks over a large number of documents (e.g., the examples shown in Fig. 2). This capability allows for quick iteration on extraction task, allowing scientists to efficiently experiment with what information it might be helpful to extract at scale from the literature. It skirts mechanics’ traditional resistance to schematization by allowing the use of ad hoc schemas designed for specific purposes, such as identifying a corpus of papers studying one particular phenomenon (e.g., elastic bistability and non-cuttability) across different contexts, and extracting parallel information from papers to study in a systematic way.

**Normalizing Representations.** In applying machine learning to mechanics, language models can supplement or replace the need for traditional feature engineering by projecting text sequences to fixed-width vector representations which encode useful meaning. LLMs can just as readily encode the phrase “hard ceramic spheres in a highly compressible aluminum foam matrix” [7] as they can any other structural composite design description, regardless of the number of the words in the description or their meaning. Combined with zero- or few-shot information extraction (discussed above) for annotating training data, these representations can then be used in downstream machine learning models, such as predicting the properties of composites from descriptions of their structure. This capability is powerful in allowing the exploitation of the kinds of ad hoc schemas described above, for concrete tasks such as property or application prediction.

**Assisted Programming.** VLLMs such as PaLM and GPT-3/4 are pretrained on large corpora of code in addition to natural language text, allowing them to generate code in response to prompts. GitHub CoPilot [57] is perhaps the most salient example of this ability. This capability is likely to have major impact in mechanics, as both the computational and experimental mechanics communities rely on algorithms in their research pipelines. Potential examples include finite element subroutines [58–60], voxel-based mechanics models [61], and algorithms for experimental characterization [62,63]. For these tasks and others, LLMs can readily develop new code and translate between programming languages.

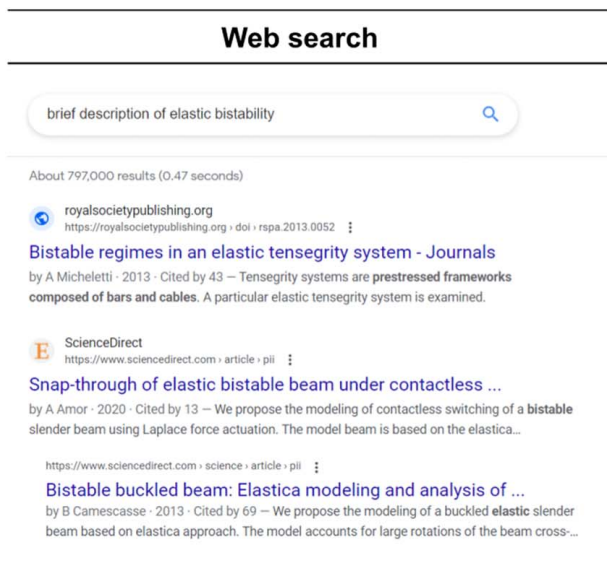
**Knowledge Recall.** VLLMs such as GPT-4 and PaLM encode a large amount of memorized knowledge in their parameters and can be used as “soft knowledge bases.” Furthermore, they can synthesize information across their memorized training data, allowing quick answers to complex questions, where a traditional search engine could only return relevant documents that would have to

be further parsed to get the desired answer. For example, a query asking to describe elastic bistability succinctly (and correctly) synthesizes information across multiple sources (Fig. 4). In theory, this allows quick access to answers which are otherwise diffuse or difficult to locate in the mechanics literature. However, as we discuss in the next sections, there are limits to LLMs' potential for factuality.

**Hypothesis Generation.** VLLMs' generative abilities can extend beyond simple question-answering, into more interpretive questions such as interesting hypotheses to explore. By synthesizing and summarizing information across large swathes of their training corpora, LLMs are able to coherently answer questions such as "what direction should I explore for..." or "what experiment should I perform to find out..." This capability can potentially allow LLMs to serve as engines for creativity in addition to knowledge and comprehension, both in mechanics and in science generally [64].

### What Limitations and Open Questions Remain for Large Language Models in Mechanics?

There are key limitations of large language models which need to be overcome before they can achieve their full potential as engines for speeding progress in mechanics and physical sciences.



### GPT-4 prompt

Briefly describe elastic bistability.

Elastic bistability is a phenomenon where an elastic structure or material exhibit two distinct stable states or configurations under the influence of an external force or load. The structure can transition between these states when a threshold force is applied, and it remains stable in either state without continuous external loading. Elastic bistability is commonly found in mechanical systems, such as switches, sensors, and actuators, and it enables energy-efficient design and control in various applications.

Fig. 4 Example of querying traditional search engine versus VLLM around a description of elastic bistability

**Hallucination.** Hallucination occurs when a language model generates information that is linguistically probable rather than factually accurate. This tendency threatens any use of LLMs as knowledge bases, and cuts into their advantage over more traditional information retrieval engines. One approach to mitigating this issue is retrieval-augmented generation [12,65], which first retrieves documents related to a query, and then prompts the LLM to generate text based on those documents. This approach is one opportunity for hybrid approaches using both MLLMs and VLLMs, as shown in Fig. 5, where database entries could be populated such as those in Refs. [66,67]. A mechanics-specific MLLM can be trained, applied across its entire training corpus, and used to retrieve relevant documents in a neural retrieval arrangement [68], and then a VLLM used to synthesize these results into a coherent response (e.g., WebGPT [69]).

**Multi-step Reasoning Errors.** Reasoning and comprehension errors threaten the use of LLMs as information extraction agents, hindering their ability to perform complex and useful comprehension across large corpora of mechanics literature. While the source of such errors is the same as hallucination (linguistic plausibility over factual/logical accuracy), the solution space is somewhat different. Extensions of few-shot learning such as chain-of-thought prompting [11] and faithful chain-of-thought prompting [70] encourage the model to generate intermediate reasoning steps, and have been found to improve the overall accuracy in performing complex reasoning tasks. Within mechanics specifically, the open questions include what comprehension and extraction tasks are useful to perform, and what strategies are needed for models to perform them reliably.

**Interpretability.** A key factor in the application of LLMs is interpretability. When a researcher queries a LLM for a scientific fact, uses it to suggest a hypothesis, or uses it to extract knowledge from a text, it is important to understand the source(s) of the model's response. This is important not just from a scientific ethics perspective (properly citing the sources of used information) but from an even-more-basic verification perspective. No fact or suggestion generated by a LLM can be verified as true without some understanding of the underlying source for that information. Given the propensity of LLMs to hallucinate, this lack of verifiability brings with it serious risks of propagating error through work that uses LLMs as knowledge sources.

Both retrieval-augmented generation and chain-of-thought prompting represent means for improving model interpretability (and thus verifiability). Retrieval-augmented generation produces the textual sources underlying the model's generated information, while chain-of-thought prompting exposes a facsimile of the model's intermediate reasoning, allowing potentially easier resolution of logical mistakes.

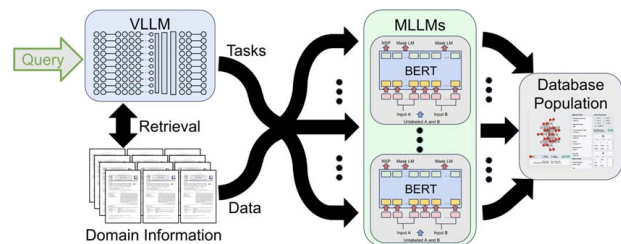


Fig. 5 Example of a hybrid approach leveraging both medium and large language models to perform database population/retrieval from the scientific literature (using Refs. [66,67] as examples). A query is issued to the VLLM, which references/retrieves domain information and issues tasks to MLLMs to extract text/values to populate the database. This framework is roughly based on Ref. [16].

**Lack of Originality.** When asked interpretive or speculative questions, VLLMs tend to produce coherent-but-uninteresting output [9], a result yet again of these models' focus on producing the most likely rather than the most interesting possible text. This compromises their ability to serve as engines for scientific creativity by suggesting hypotheses or promising research directions. The increasing power of LLMs hints at their potential for helping to automate hypothesis discovery, but generating detailed scientific ideas that are both novel and useful remains an open problem [64]. While it is a less-studied problem than those previously discussed, there has been work on using scientific NLP to inculcate creativity in science, often focused on analogy as a useful construct [71–73].

**Outlooks for Large Language Models in Mechanics.** In the previous section, we discuss several key capabilities of LLMs and speculate about how they might apply to the field of mechanics. However, one of the biggest open questions in applying LLMs to mechanics is which of these capabilities are likely to produce the most effective results in terms of the overall goals of the field. Although the information extraction capabilities of VLLMs offer unprecedented ability to perform complex reading comprehension tasks at scale, higher impact results may arise from using VLLMs for hypothesis generation, or otherwise. The answers to this big question will decide the role of LLMs in applied mechanics going forward, and these answers will remain unknown until mechanics-based LLMs are actually built and applied.

## Acknowledgment

NRB and SHD acknowledge support from NSWC Grant #N00174-22-1-0020. CM acknowledges support from the NASA Space Technology Graduate Research Opportunities Fellowship (Grant No. 80NSSC19K1164). DD acknowledges support from NSF Grant 2033558. TMP and MPE acknowledge the support of a Department of Defense Vannevar Bush Fellowship, Grant ONR N00014-18-1-3031.

## Conflict of Interest

There are no conflicts of interest.

## Data Availability Statement

No data, models, or code were generated or used for this paper.

## References

- [1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I., 2017, "Attention is All You Need," *Adv. Neural Inf. Process. Syst.*, **30**.
- [2] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K., 2018, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv:1810.04805.
- [3] Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., et al., 2022, "PaLM: Scaling Language Modeling With Pathways," arXiv:2204.02311.
- [4] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., and Neelakantan, A., et al., 2020, "Language Models Are Few-Shot Learners," *Adv. Neural Inf. Process. Syst.*, **33**, pp. 1877–1901.
- [5] OpenAI, 2023, "GPT-4 Technical Report," arXiv:2303.08774.
- [6] OpenAI, 2023, "Aligning Language Models to Follow Instructions."
- [7] Szyniszewski, S., Vogel, R., Bittner, F., Jakubczyk, E., Anderson, M., Pelacci, M., Chinedu, A., Endres, H.-J., and Hipke, T., 2020, "Non-cutttable Material Created Through Local Resonance and Strain Rate Effects," *Sci. Rep.*, **10**(1), p. 11539.
- [8] Cresswell-Boyes, A. J., Davis, G. R., Krishnamoorthy, M., Mills, D., and Barber, A. H., 2022, "Composite 3D Printing of Biomimetic Human Teeth," *Sci. Rep.*, **12**(1), p. 7830.
- [9] Lahat, A., Shachar, E., Avidan, B., Shatz, Z., Glicksberg, B. S., and Klang, E., 2023, "Evaluating the Use of Large Language Model in Identifying Top Research Questions in Gastroenterology," *Sci. Rep.*, **13**, p. 4164.
- [10] Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., and Neubig, G., 2023, "Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing," *ACM Comput. Surv.*, **55**(9), pp. 195:1–195:35.
- [11] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., and Zhou, D., 2023, "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models," arXiv:2201.11903.
- [12] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., et al., 2020, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *Adv. Neural Inf. Process. Syst.*, **33**, pp. 9459–9474.
- [13] Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Zettlemoyer, L., Cancedda, N., and Scialom, T., 2023, "Toolformer: Language Models Can Teach Themselves to Use Tools," arXiv:2302.04761.
- [14] Mialon, G., Dessì, R., Lomeli, M., Nalmpantis, C., Pasunuru, R., Raileanu, R., Rozière, B., et al., 2023, "Augmented Language Models: A Survey," arXiv:2302.07842.
- [15] Safavi, T., Downey, D., and Hope, T., 2022, "CascadER: Cross-Modal Cascading for Knowledge Graph Link Prediction," *Tech. Rep.*, arXiv:2205.08012.
- [16] Ge, Y., Hua, W., Ji, J., Tan, J., Xu, S., and Zhang, Y., 2023, "OpenAGI: When LLM Meets Domain Experts," arXiv:2304.04370.
- [17] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J., 2020, "Exploring the Limits of Transfer Learning With a Unified Text-to-Text Transformer," *J. Mach. Learn. Res.* **21**(1), pp. 5485–5551.
- [18] Taylor, R., Kardas, M., Cucurull, G., Scialom, T., Hartshorn, A., Saravia, E., Poulton, A., Kerkez, V., and Stojnic, R., 2022, Galactica: A Large Language Model for Science," arXiv:2211.09085.
- [19] Beltagy, I., Lo, K., and Cohan, A., 2019, "SciBERT: A Pretrained Language Model for Scientific Text," arXiv:1903.10676.
- [20] Hong, Z., Ajith, A., Pauloski, G., Duede, E., Malamud, C., Magoulas, R., Chard, K., and Foster, I., 2022, "ScholarBERT: Bigger Is Not Always Better," 2022arXiv:220511342H.
- [21] Trewartha, A., Walker, N., Huo, H., Lee, S., Cruse, K., Dagdelen, J., Dunn, A., Persson, K. A., Ceder, G., and Jain, A., 2022, "Quantifying the Advantage of Domain-Specific Pre-training on Named Entity Recognition Tasks in Materials Science," *Patterns*, **3**(4), p. 100488.
- [22] Gupta, T., Zaki, M., Krishnan, N. M. A., and Mausam, 2022, "MatSciBERT: A Materials Domain Language Model for Text Mining and Information Extraction," *npj Comput. Mater.*, **8**(1), pp. 1–11.
- [23] Yoshitake, M., Sato, F., Kawano, H., and Teraoka, H., 2022, "MaterialBERT for Natural Language Processing of Materials Science Texts," *Sci. Technol. Adv. Mater.: Methods*, **2**(1), pp. 372–380.
- [24] Huang, S., and Cole, J. M., 2022, "BatteryBERT: A Pretrained Language Model for Battery Database Enhancement," *J. Chem. Inf. Model.*, **62**(24), pp. 6365–6367.
- [25] Zhao, J., Huang, S., and Cole, J. M., 2023, "OpticalBERT and OpticalTable-SQA: Text- and Table-Based Language Models for the Optical-Materials Domain," *J. Chem. Inf. Model.*, **63**(7), pp. 1961–1981.
- [26] Xu, C., Wang, Y., and Barati Farimani, A., 2023, "TransPolymer: A Transformer-Based Language Model for Polymer Property Predictions," *npj Comput. Mater.*, **9**(1), pp. 1–14.
- [27] Petroni, F., Rocktäschel, T., Lewis, P., Bakhtin, A., Wu, Y., Miller, A., and Riedel, S., 2019, "Language Models as Knowledge Bases?," arXiv preprint arXiv:1909.01066.
- [28] Kandpal, N., Deng, H., Roberts, A., Wallace, E., and Raffel, C., 2022, "Large Language Models Struggle to Learn Long-Tail Knowledge," arXiv:2211.08411.
- [29] Balabin, H., Hoyt, C. T., Birkenbihl, C., Gyori, B. M., Bachman, J., Kodamullil, A. T., Plöger, P. G., Hofmann-Apitius, M., and Domingo-Fernández, D., 2022, "STonKGs: A Sophisticated Transformer Trained on Biomedical Text and Knowledge Graphs," *Bioinformatics*, **38**(6), pp. 1648–1656.
- [30] Nadkarni, R., Wadden, D., Beltagy, I., Smith, N. A., Hajishirzi, H., and Hope, T., 2021, "Scientific Language Models for Biomedical Knowledge Base Completion: An Empirical Study," arXiv:2106.09700.
- [31] Naik, A., Parasa, S., Feldman, S., Wang, L. L., and Hope, T., 2021, "Literature-Augmented Clinical Outcome Prediction," ArXiv.
- [32] Mikolov, T., Chen, K., Corrado, G., and Dean, J., 2013, "Efficient Estimation of Word Representations in Vector Space," arXiv:1301.3781.
- [33] Pennington, J., Socher, R., and Manning, C. D., 2014, "GloVe: Global Vectors for Word Representation," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, Oct. 25–29, pp. 1532–1543.
- [34] Tshitoyan, V., Dagdelen, J., Weston, L., Dunn, A., Rong, Z., Kononova, O., Persson, K. A., Ceder, G., and Jain, A., 2019, "Unsupervised Word Embeddings Capture Latent Knowledge From Materials Science Literature," *Nature*, **571**(7763), pp. 95–98.
- [35] Shetty, P., and Ramprasad, R., 2021, "Automated Knowledge Extraction From Polymer Literature Using Natural Language Processing," *iScience*, **24**(1), p. 101922.
- [36] Pei, Z., Yin, J., Liaw, P. K., and Raabe, D., 2023, "Toward the Design of Ultrahigh-Entropy Alloys Via Mining Six Million Texts," *Nat. Commun.*, **14**(1), p. 54.
- [37] Yang, F., 2021, "Natural Language Processing Applied on Large Scale Data Extraction From Scientific Papers in Fuel Cells," 2021 5th International Conference on Natural Language Processing and Information Retrieval (NLPPIR), NLPPIR 2021, Sanya, China, Dec. 17–20, Association for Computing Machinery, pp. 168–175.
- [38] Stokel-Walker, C., and Van Noorden, R., 2023, "What ChatGPT and Generative AI Mean for Science," *Nature*, **614**(7947), pp. 214–216.

- [39] Morris, M. R., 2023, "Scientists' Perspectives on the Potential for Generative AI in Their Fields," arXiv:2304.01420.
- [40] Dunn, A., Dagdelen, J., Walker, N., Lee, S., Rosen, A. S., Ceder, G., Persson, K., and Jain, A., 2022, "Structured Information Extraction From Complex Scientific Text With Fine-Tuned Large Language Models," arXiv:2212.05238.
- [41] Xie, T., Wan, Y., Huang, W., Zhou, Y., Liu, Y., Linghu, Q., Wang, S., Kit, C., Grazian, C., Zhang, W., and Hoex, B., 2023, "Large Language Models as Master Key: Unlocking the Secrets of Materials Science With GPT," arXiv:2304.02213.
- [42] Jablonka, K. M., Schwaller, P., Ortega-Guerrero, A., and Smit, B., 2023, "Is GPT-3 All You Need for Low-Data Discovery in Chemistry?," *ChemRxiv*.
- [43] Polak, M. P., and Morgan, D., 2023, "Extracting Accurate Materials Data From Research Papers With Conversational Language Models and Prompt Engineering—Example of ChatGPT," arXiv:2303.05352.
- [44] Hu, Y., and Buehler, M. J., 2023, "Deep Language Models for Interpretative and Predictive Materials Science," *APL Mach. Learn.*, **1**(1), p. 010901.
- [45] Buehler, M. J., 2022, "Modeling Atomistic Dynamic Fracture Mechanisms Using a Progressive Transformer Diffusion Model," *ASME J. Appl. Mech.*, **89**(12), p. 121009.
- [46] Buehler, M. J., 2022, "Multiscale Modeling at the Interface of Molecular Mechanics and Natural Language Through Attention Neural Networks," *Acc. Chem. Res.*, **55**(23), pp. 3387–3403.
- [47] Luu, R. K., and Buehler, M. J., 2023, "Materials Informatics Tools in the Context of Bio-inspired Material Mechanics," *ASME J. Appl. Mech.*, **90**(9), p. 090801.
- [48] Henderson, M. R., and Taylor, L. E., 1993, "A Meta-model for Mechanical Products Based Upon the Mechanical Design Process," *Res. Eng. Des.*, **5**(3), pp. 140–160.
- [49] Nayak, A., Geleda, B., Sakhapara, A., Singh, A., and Acharya, N., 2015, "Visualization of Mechanics Problems Based on Natural Language Processing," *Int. J. Comput. Appl.*, **116**(14), pp. 34–37.
- [50] Shi, F., Chen, L., Han, J., and Childs, P., 2017, "A Data-Driven Text Mining and Semantic Network Analysis for Design Information Retrieval," *ASME J. Mech. Des.*, **139**(11), p. 111402.
- [51] Martinez-Gil, J., Freudenthaler, B., and Natschläger, T., 2017, "Automatic Recommendation of Prognosis Measures for Mechanical Components Based on Massive Text Mining," Proceedings of the 19th International Conference on Information Integration and Web-Based Applications & Services, iiWAS '17, Salzburg, Austria, Dec. 4–6, Association for Computing Machinery, pp. 32–39.
- [52] Lee, H. S., Song, H. G., and Lee, H. S., 2013, "Classification of Photovoltaic Research Papers by Using Text-Mining Techniques," *Appl. Mech. Mater.*, **284–287**, pp. 3362–3369.
- [53] Zhang, J., Yuan, J., and Xu, J., 2022, "An Artificial Intelligence Technology Based Algorithm for Solving Mechanics Problems," *IEEE Access*, **10**, pp. 92971–92985.
- [54] Dai, J., Lu, L., Leanza, S., Hutchinson, J., and Zhao, R. R., 2023, "Curved Ring Origami: Bistable Elastic Folding for Magic Pattern Reconfigurations," *J. Appl. Mech.*, pp. 1–27.
- [55] Wan, G., Liu, Y., Xu, Z., Jin, C., Dong, L., Han, X., Zhang, J. X. J., and Chen, Z., 2020, "Tunable Bistability of a Clamped Elastic Beam," *Ext. Mech. Lett.*, **34**, p. 100603.
- [56] Rafsanjani, A., and Pasini, D., 2016, "Bistable Auxetic Mechanical Metamaterials Inspired by Ancient Geometric Motifs," *Ext. Mech. Lett.*, **9**, pp. 291–296.
- [57] Github Copilot.
- [58] Smith, M., 2009, ABAQUS/Standard User's Manual, Version 6.9. Dassault Systèmes Simulia Corp, Providence, RI.
- [59] Dawson, P. R., and Boyce, D. E., 2015, "FEpX—Finite Element Polycrystals: Theory, Finite Element Formulation, Numerical Implementation and Illustrative Examples," arXiv:1504.03296.
- [60] Quey, R., and Kasemer, M., 2022, "The Neper/FEpX Project: Free/Open-Source Polycrystal Generation, Deformation Simulation, and Post-Processing," *IOP Conf. Ser.: Mater. Sci. Eng.*, **1249**(1), p. 012021.
- [61] Lebensohn, R. A., Kanjarla, A. K., and Eisenlohr, P., 2012, "An Elasto-viscoplastic Formulation Based on Fast Fourier Transforms for the Prediction of Micromechanical Fields in Polycrystalline Materials," *Int. J. Plast.*, **32–33**, pp. 59–69.
- [62] DeGraef, M., Jackson, M., Kleingers, J., Zhu, C., Tessmer, J., Lenthe, W. C., Singh, S., Atkinson, M., Wright, S., and Ánes, H., 2019, "EMsoft-org/EMsoft: EMsoft Release 5.0.0," *Zenodo*.
- [63] Callahan, P. G., and DeGraef, M., 2013, "Dynamical Electron Backscatter Diffraction Patterns. Part I: Pattern Simulations," *Microsc. Microanal.*, **19**(5), pp. 1255–1265.
- [64] Hope, T., Downey, D., Etzioni, O., Weld, D. S., and Horvitz, E., 2022, "A Computational Inflection for Scientific Discovery," Tech. Rep., arXiv:2205.02007.
- [65] Izcard, G., Lewis, P., Lomeli, M., Hosseini, L., Petroni, F., Schick, T., Dwivedi-Yu, J., Joulin, A., Riedel, S., and Grave, E., 2022, "Atlas: Few-Shot Learning With Retrieval Augmented Language Models," arXiv:2208.03299.
- [66] Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., and Persson, K. A., 2013, "Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation," *APL Mater.*, **1**(1), p. 011002.
- [67] Borg, C. K. H., Frey, C., Moh, J., Pollock, T. M., Gorsse, S., Miracle, D. B., Senkov, O. N., Meredig, B., and Saal, J. E., 2020, "Expanded Dataset of Mechanical Properties and Observed Phases of Multi-principal Element Alloys," *Sci. Data*, **7**(1), p. 430.
- [68] Mitra, B., and Craswell, N., 2018, "An Introduction to Neural Information Retrieval," Tech. Rep.
- [69] Nakano, R., Hilton, J., Balaji, S., Wu, J., Ouyang, L., Kim, C., Hesse, C., et al., 2022, "WebGPT: Browser-Assisted Question-Answering With Human Feedback," arXiv:2112.09332.
- [70] Lyu, Q., Havaldar, S., Stein, A., Zhang, L., Rao, D., Wong, E., Apidianaki, M., and Callison-Burch, C., 2023, "Faithful Chain-of-Thought Reasoning," arXiv:2301.13379.
- [71] Hope, T., Chan, J., Kittur, A., and Shahaf, D., 2017, "Accelerating Innovation Through Analogy Mining," *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 235–243.
- [72] Lahav, D., Falcon, J. S., Kuehl, B., Johnson, S., Parasa, S., Shomron, N., Chau, D. H., et al., 2022, "A Search Engine for Discovery of Scientific Challenges and Directions," *Proc. AAAI Conf. Artif. Intell.*, **36**(11), pp. 11982–11990.
- [73] Kang, H. B., Qian, X., Hope, T., Shahaf, D., Chan, J., and Kittur, A., 2022, "Augmenting Scientific Creativity With an Analogical Search Engine," *ACM Trans. Comput.-Hum. Interact.*, **29**(6), pp. 1–36.