Empowering Materials Science with Large Language Models: An Overview of MatCopilot®

Yu-Yang Liu[a], **Chuan-Wei Kuo**[a]

[a] *DeepVerse, Shanghai, China* ceo@deepverse.tech, auberonkuo@deepverse.tech

* Presenting author

1. Introduction

Over the past two decades, the rapid evolution of artificial intelligence (AI) has profoundly impacted both industry and academia. Early breakthroughs in deep neural networks, unsupervised pre-training, and attention-based architectures have led to today's era of large language models (LLMs). Modern LLMs demonstrate remarkable capabilities in natural language understanding, few-shot learning, and generative tasks. Concurrently, AI for Science—particularly in materials and life sciences—has made unprecedented progress, epitomized by models that accurately predict protein structures or generate new candidate materials.

In materials science, researchers increasingly rely on advanced machine learning methods, including graph neural networks (GNNs) and generative approaches such as diffusion models, to handle complex tasks like property prediction and new compound discovery. These techniques address key challenges in high-dimensional, sparse, and noisy materials data, offering opportunities to drastically accelerate innovation cycles.

However, bridging LLM capabilities with domain-specific complexities remains a non-trivial undertaking. Training and fine-tuning large models in specialized fields involves not only compressing and distilling extensive knowledge but also integrating physical constraints to ensure accuracy and reliability. Recent open-source and commercial models alike illustrate the potential of collaborative human–machine workflows: domain experts guide an LLM's reasoning, while the LLM proposes new hypotheses or experimental designs to expedite materials discovery.

2. MatCopilot®: AI-Driven Materials Discovery Platform

DeepVerse is committed to enabling efficient innovation, reducing research risks, and lowering development costs in materials science. Our flagship product, **MatCopilot®**, is a revolutionary AI platform tailored for researchers and engineers in this domain. It integrates multiple cutting-edge techniques:

- **Physics-Informed Neural Networks:** Incorporates fundamental physical laws to guide predictive modeling and reduce errors.
- **Uncertainty Quantification and Active Learning:** Improves predictive reliability by sampling the most informative data points and iteratively refining the model.
- **Specialized LLM Compression:** Employs a proprietary "diffusive chain-of-thought" framework and hybrid compression strategies to distill large language models into a domain-specific expert LLM, focusing on materials science tasks.

By combining data-driven insights with theoretical knowledge, MatCopilot® efficiently handles high-dimensional and noisy datasets. The system leverages an extensive knowledge graph covering materials-specific literature, enabling rapid exploration of complex search spaces. Built to accelerate discovery across various industry verticals—from battery materials and sustainable chemistry to high-performance alloys—MatCopilot® has already demonstrated success through several pilot projects.

Key Capabilities

- *Accelerated Iterations:* Automates experiment planning and hypothesis generation, guided by real-time feedback loops.
- *Customized Solutions:* Provides bespoke modeling and analytics pipelines to address specific R&D projects, helping partners reduce time-to-market.
- *Scalable Expertise:* Offers large-scale knowledge integration, bridging engineering requirements, scientific literature, and advanced AI algorithms for materials discovery.

Conclusion

Large language models, coupled with robust domain knowledge and physics-informed approaches, are reshaping the landscape of materials R&D. Tools like MatCopilot® streamline the traditionally time-consuming cycle of hypothesis generation, simulation, validation, and refinement. As AI continues to advance, it promises not only to accelerate innovation but also to democratize the discovery process, enabling teams of all scales to tackle increasingly complex materials challenges.