

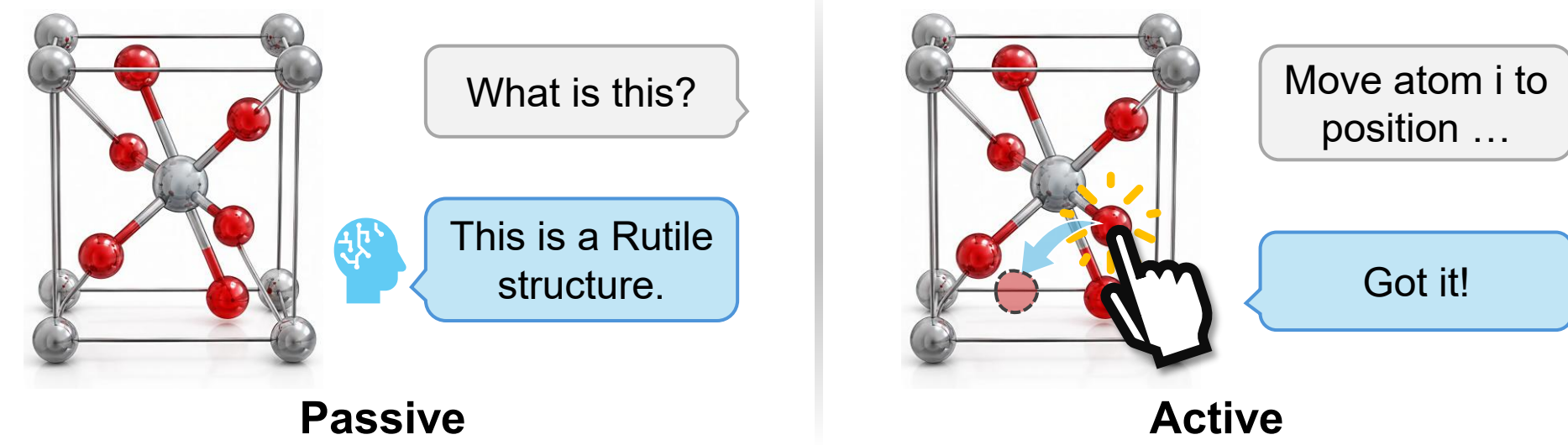
AtomWorld: Beyond Structure Recognition

Evaluating Spatial Reasoning Through Atomic Manipulation

Taoyuze Lv^{1,*}, Alexander Chen^{2,*}, Fengyu Xie^{1,#}, Chu Wu¹, Jeffrey Meng², Dongzhan Zhou³, Bram Hoex², Yingheng Wang⁴, Zhicheng Zhong¹, Tong Xie^{2,5,#}

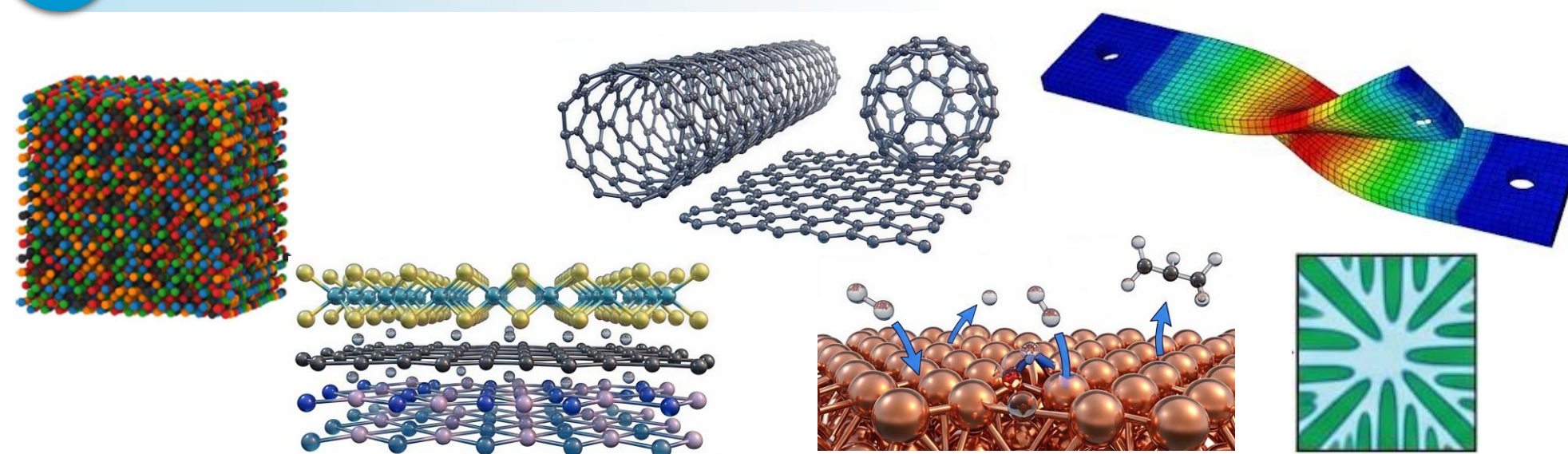
Why Do We Need AtomWorld?

1 Static tasks are not enough



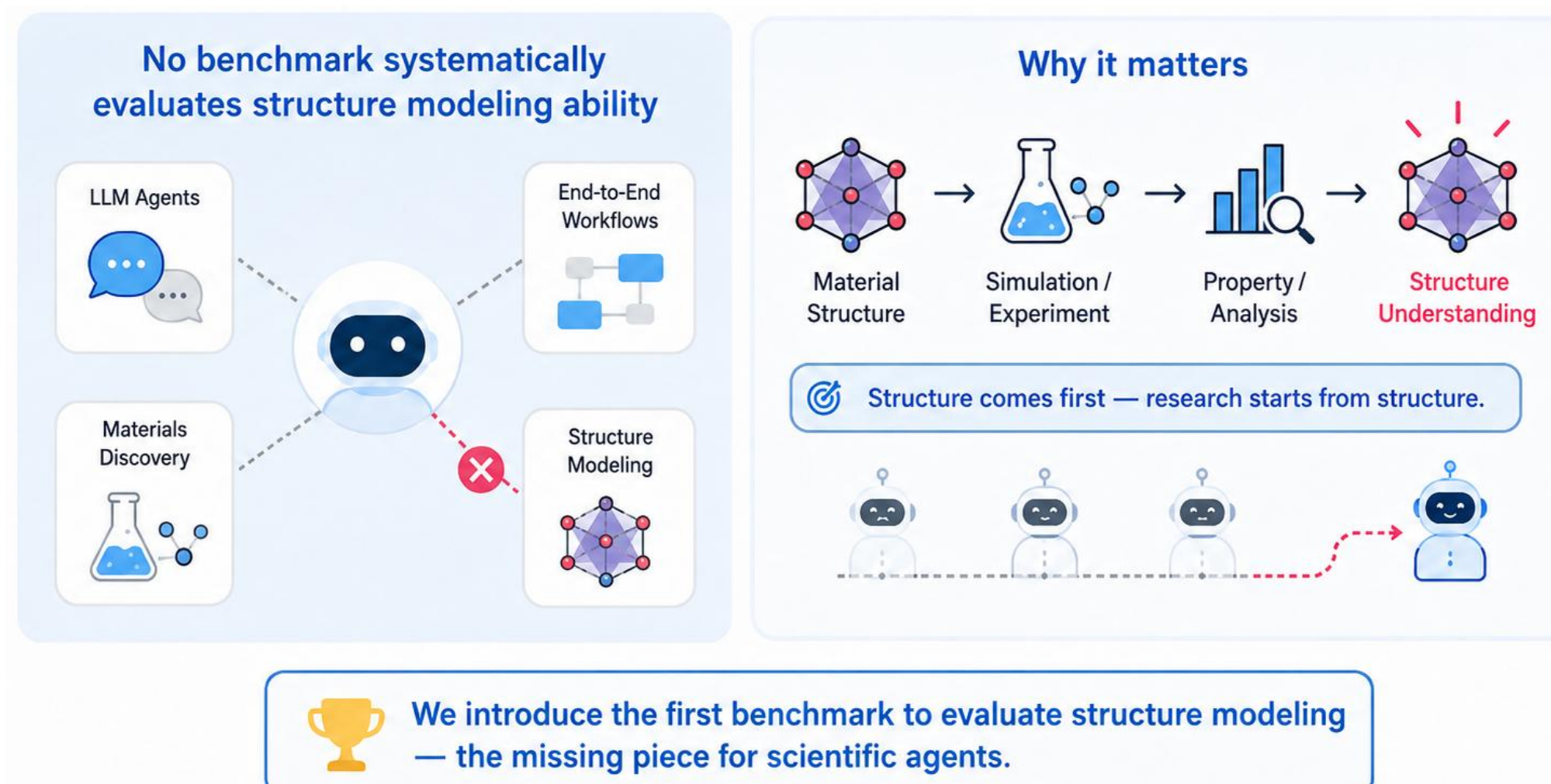
We evaluate the ability to act, not just to describe. Also, currently reasoning for manipulating is lacking. Reasoning is more about knowing than doing.

2 Modeling matters in science

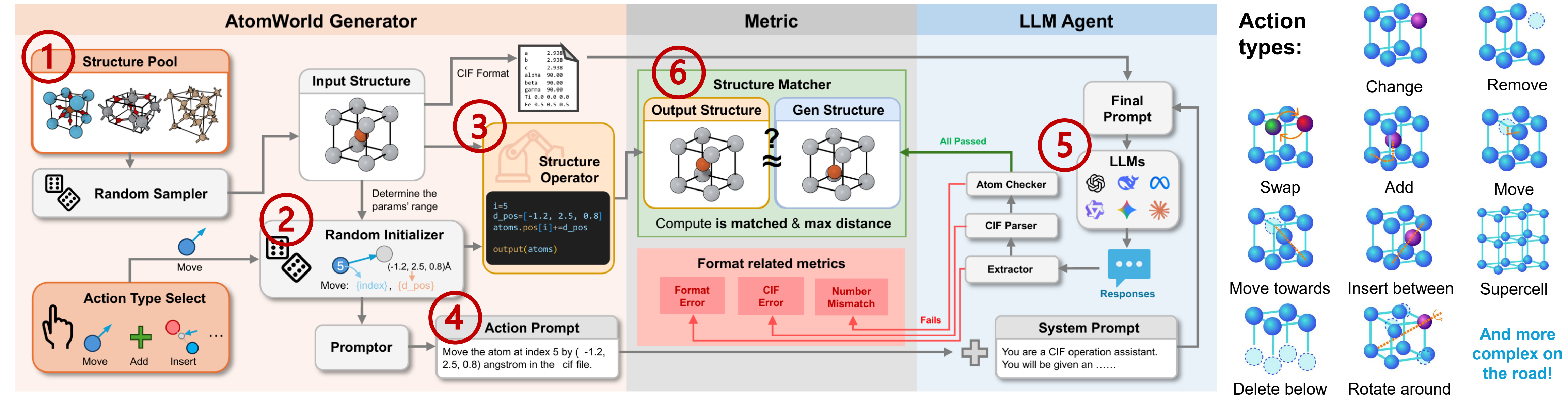


Real modeling workflows are iterative, open-ended, and construction-oriented. The modelling is not a fixed pipeline, but an unbounded design space. Combinatorial possibilities are almost infinite.

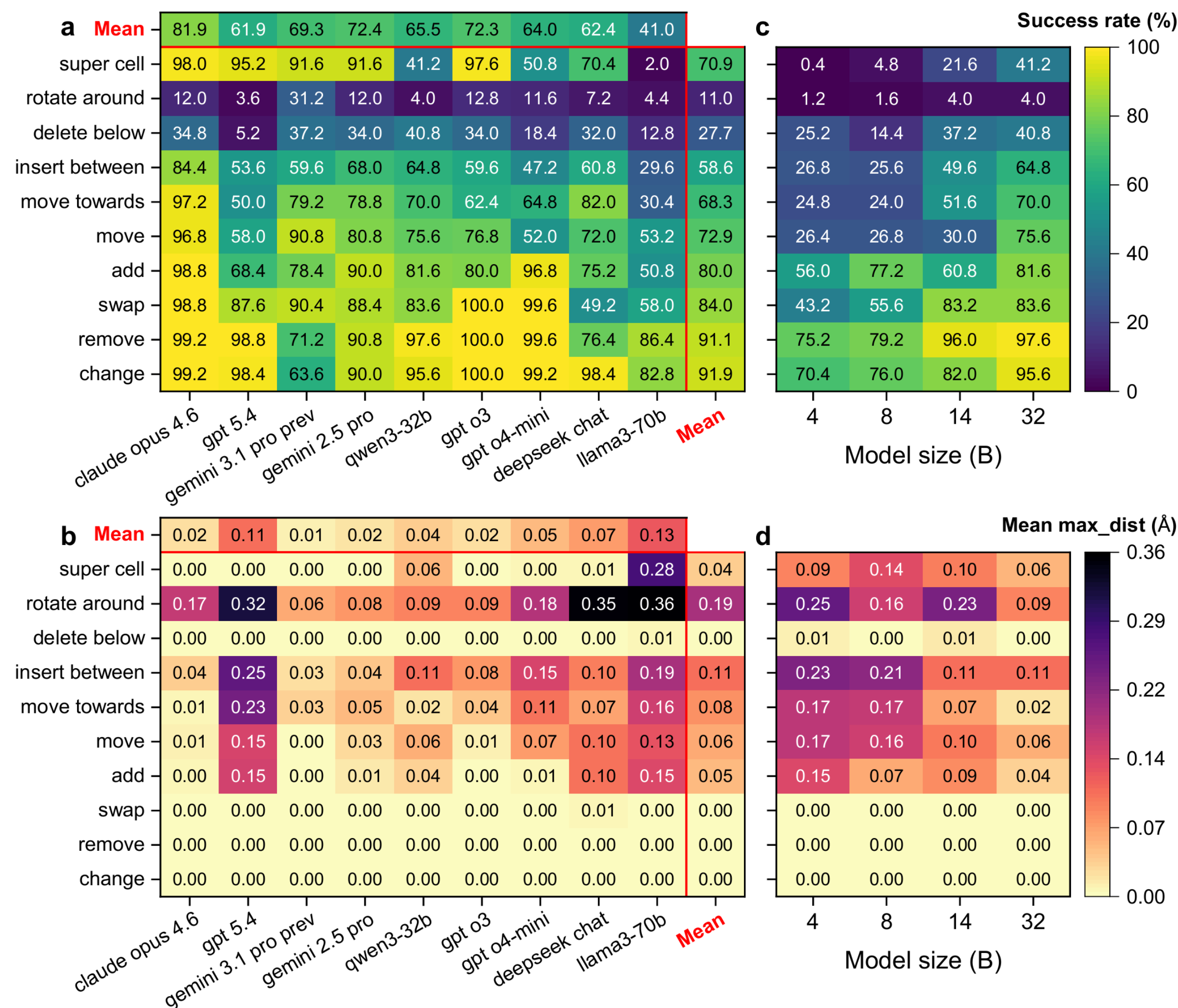
3 One missing piece for scientific agents



Benchmark Design



Do LLMs Understand Modeling?



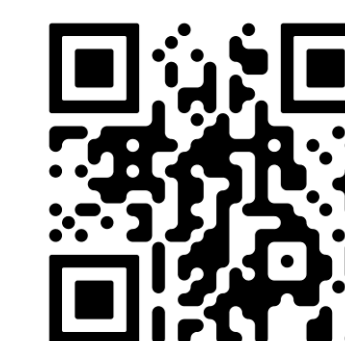
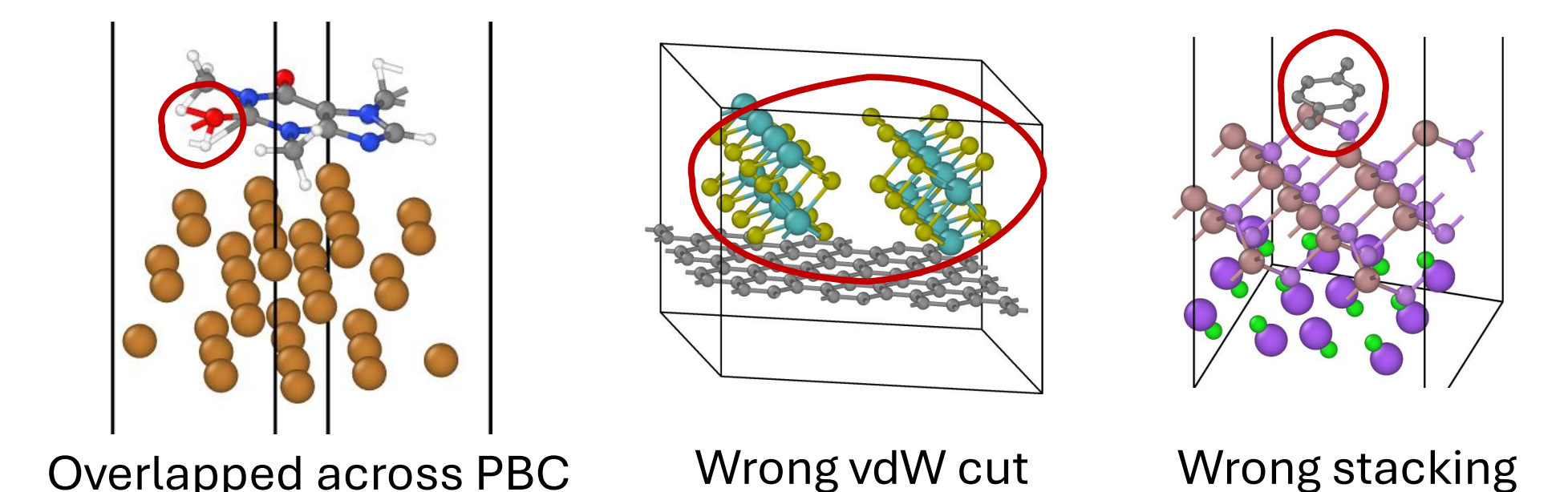
Main Takeaways

Scientific intelligence requires action, not just understanding.

Current LLMs can struggle to construct, modify, and verify them. Bridging this gap may require action-driven learning, richer feedback, and interactive environments.

Beyond the Benchmark:

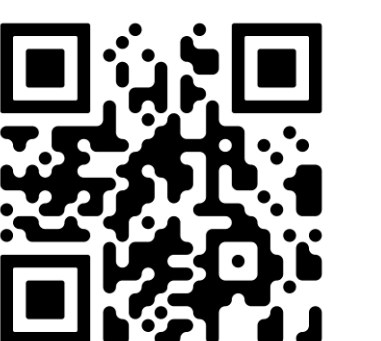
In practical attempts to build agentic systems, we found that providing LLMs with domain skills and coding tools does not reliably make agents good at modeling:



GitHub Repo



Leaderboard



Hugging Face