

nanoRA: A Multi-Agent Research Assistant for Autonomous Scientific Workflows on nanoHUB

Juan C. Verduzco, Mohnish Harwani, and Alejandro Strachan

MOTIVATION

Large language models (LLMs) are playing an increasingly important role in science and engineering with applications in code generation¹, autonomous experimentation^{2,3}, and literature summarization⁴.

Agentic AI systems such as AtomAgents² (atomistic simulations), Coscientist³ (cloud lab automation), ChemCrow⁵ (chemical synthesis), and EI Agente⁶ (quantum chemistry) have shown how LLMs can plan, execute, and adapt scientific workflows with minimal human input.

Despite early successes, adoption remains limited by rigid tool integrations and domain-specific workflows that lack generalizable, scalable execution frameworks.

METHODS

nanoRA is an agentic AI system that leverages a team of independent agents powered by an LLM (GPT-4o) to solve scientific endeavors.



I am the Supervisor agent !
I interpret the user's request, plan the workflow, and assign tasks to the correct agent. I ensure each step runs smoothly, data flows correctly, and the final response is complete and helpful.



I am the nanoHUB agent !
I can connect to tools within the nanoHUB infrastructure. I can search for tools, retrieve metadata, query results, and run simulations with user approval.



I am the Code Execution agent !
I write and run Python scripts for data analysis and visualization. I work with structured data from other agents to compute results, fit models, and generate plots.



I am the Materials Project agent!
I am responsible for retrieving materials information and atomistic structures from the Materials Project.

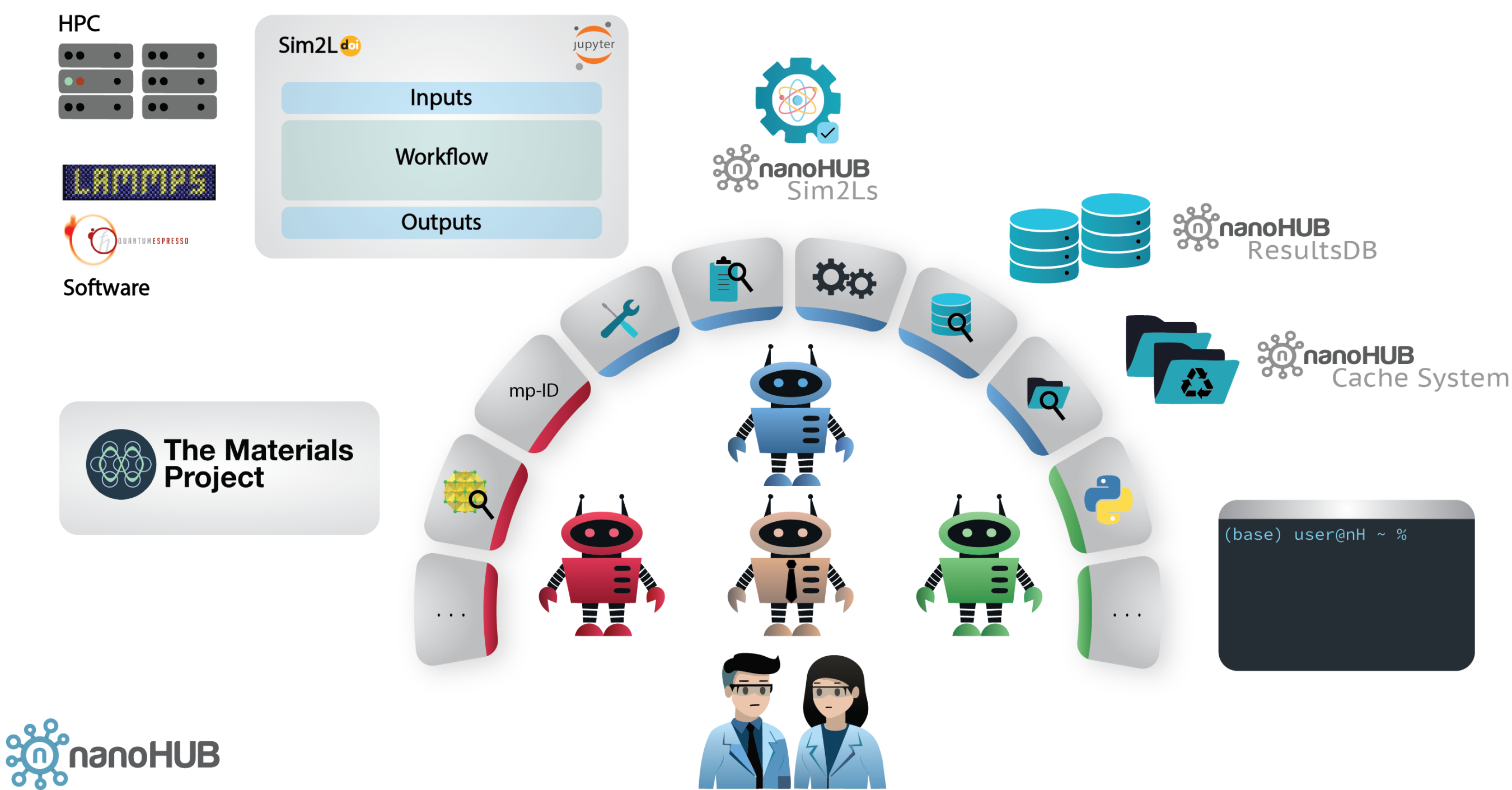
Each agent has its own tools to plan, select, and run scientific workflows.

nanoRA is powered by **Sim2Ls**, standardized end-to-end simulation workflows with validated metadata.




FRAMEWORK

nanoRA separates reasoning and orchestration from execution by interacting with nanoHUB's Sim2L API and external data sources, enabling reproducible, multi-step research tasks.




ALPHAFOLD

nanoRA was tasked with specific scientific objectives. For example, predicting protein sequences, querying pressure - induced phase transitions, and run simple MD simulations.




Predict the shape of the following protein sequence: "NLVIQWLKDGPPSSGRPPPS".




Can you predict the shape of this protein?

No. I can help you find a tool in nanoHUB, though.




Find me some tools that specialize in protein structure prediction.

I found two! There's AlphaFold and ESMFold.




Which tool you would like to use? AlphaFold or ESMFold.

Use AlphaFold.




Let's run AlphaFold.

Alright, running simulation.

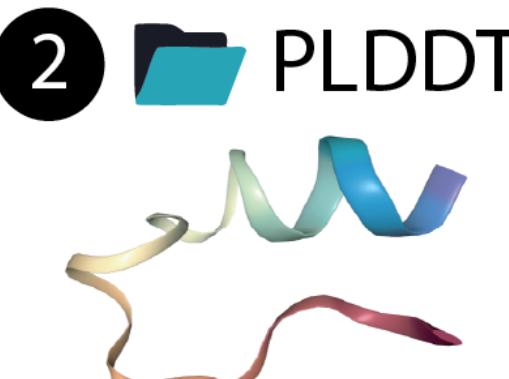


The AlphaFold simulation has been completed. Here are the results.

1 PLDDT



2 PLDDT



Tool Parameters: AlphaFold

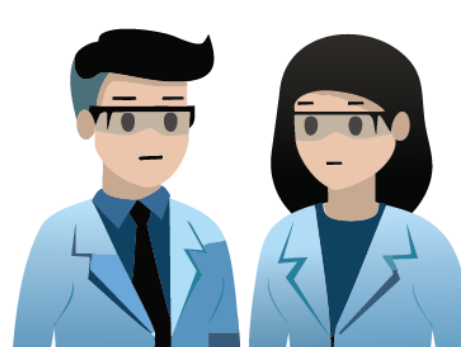
Tool Execution: AlphaFold

High Performance Computing


nanoHUB ResultsDB

nanoHUB Cache System


PHASE TRANSITION




I want to analyze phase transitions for SnC. Query 200 points from nanoHUB. Retrieve their energy, volume, pressure, and spacegroup. Plot energy versus volume, colored by spacegroup.



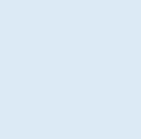
Tool Search




Tool Parameters: CellRelaxDFT



Query Database: CellRelaxDFT

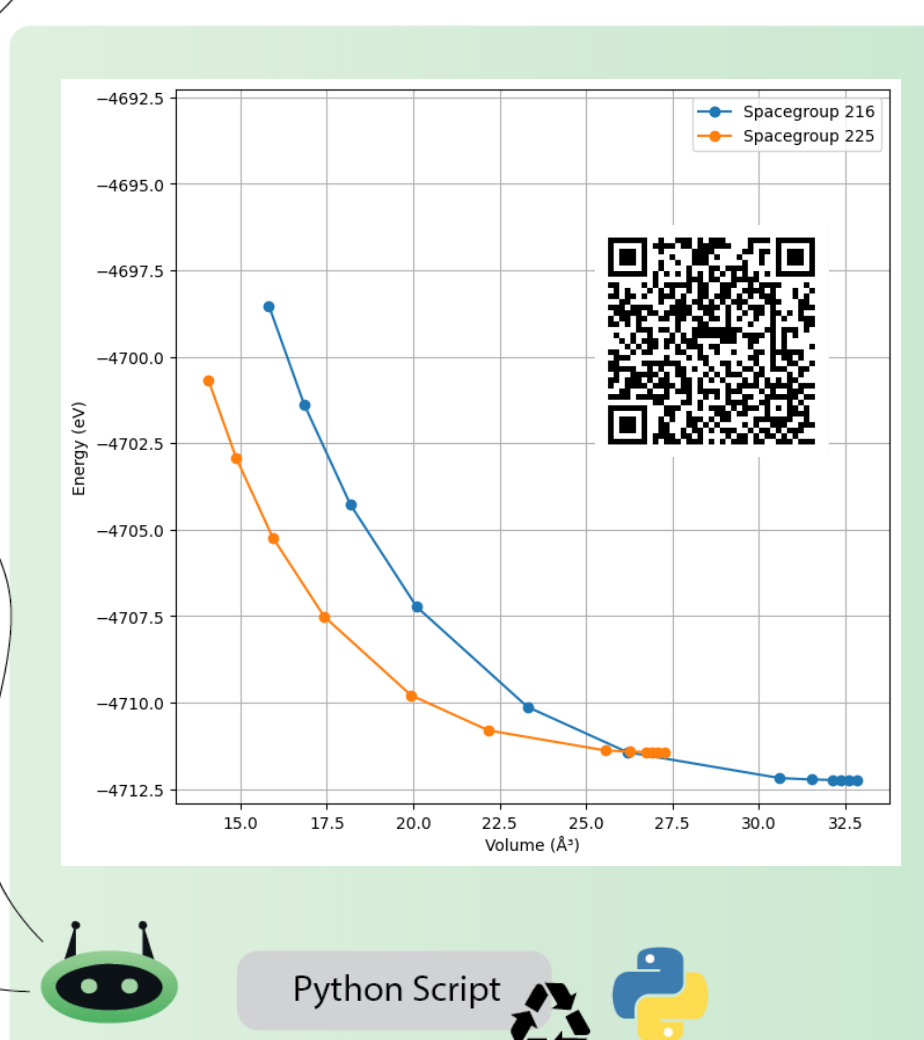


MP Query: SnC



Python Script

The analysis of phase transitions for SnC has been completed. A scatter plot of volume versus energy, color-coded by spacegroup, was generated using data from the 'cellrelaxDFT' tool on nanoHUB. This visualization will help you analyze phase transitions



CONCLUSIONS

nanoRA connects LLMs with real-world science by using modular workflows instead of hard-coded instructions, making it easier to run flexible, multi-step experiments without needing custom programming.

nanoRA runs on nanoHUB's Sim2L system, giving it access to 700+ computational workflows through a standard API, each with built-in checks, tracking, and automatic results storage.

Researchers can expand nanoRA's capabilities by adding new Sim2Ls, allowing the system to explore and interpret simulations across different fields and move towards more general-purpose scientific automation.

REFERENCES

- Lei, G., Docherty, R., & Cooper, S. J. (2024). Materials science in the era of large language models: a perspective. *Digital Discovery*, 3(7), 1257-1272.
- Ghafariollahi, A., & Buehler, M. J. (2025). Automating alloy design and discovery with physics-aware multimodal multiagent AI. *Proceedings of the National Academy of Sciences*, 122(4), e2414074122.
- Boiko, D. A., MacKnight, R., Kline, B., & Gomes, G. (2023). Autonomous chemical research with large language models. *Nature*, 624(7992), 570-578.
- Gupta, T., Zaki, M., Krishnan, N. A., & Mausam. (2022). MatSciBERT: A materials domain language model for text mining and information extraction. *npj Computational Materials*, 8(1), 102.
- Bran, A. M., Cox, S., Schiller, O., Baldassari, C., White, A. D., & Schwaller, P. (2023). Chemcrow: Augmenting large-language models with chemistry tools. *arXiv preprint arXiv:2304.05376*.
- Zou, Y., Cheng, A. H., Aldossary, A., Bai, J., Leong, S. X., Campos-Gonzalez-Angulo, J. A., ... & Aspuru-Guzik, A. (2025). El Agente: An Autonomous Agent for Quantum Chemistry. *arXiv preprint arXiv:2505.02484*.

