It Runs—but Is It Right? Ontology-Constrained Synthesis of Physics-Valid Simulation Tools for AI Agents

M. Xavier[1], D. Berglund[1*], B. Babu[1], E. Berglund[1]

[1] Aerobase Innovations AB, Kaserngatan 4, SE-974 42 Luleå, Sweden

*daniel.berglund@aerobase.se

Abstract

A simulation tool can run flawlessly and still be wrong. For AI agents that synthesize tools for other AI agents [4], this is a particularly dangerous failure: inconsistent units, incompatible material states, or use of a solver outside its validated domain can produce convincing outputs that are physically meaningless.

We present an ontology-constrained approach for PHASES [1], aligned with EMMO and OSMO [2,3], in which AI agents synthesize new application-level simulation tools from existing validated solvers and approved external physics components. Given a request such as “build a laser-hardening tool,” the agent first constructs a semantic blueprint that specifies the required quantities, units, material states, physical capabilities, solver dependencies, and operating envelopes. It then assembles a bounded workflow, namely laser energy deposition, transient heat transfer, PHASES transformation kinetics, and hardness and case-depth post-processing. The agent generates the interfaces, coupling logic, parameterization, and post-processing, but does not invent new governing physics.

Before registration, each candidate undergoes dimensional, semantic, state-transfer, and applicability checks, followed by sandboxed execution and physics-based validation. If a required capability is missing or incompatible, synthesis stops safely. An accepted tool carries a machine-readable validity manifest that records its operating limits, verification evidence, and ontology, material, and solver versions. The tool can then be reused by PHASES AI or other LLM agents through a tool registry and MCP. We will evaluate the approach on custom metallurgical simulation requests that require different combinations of PHASES and approved external solvers, measuring task success, correct capability selection, safe rejection, envelope compliance, and silent physical errors. The aim is to create AI-generated simulation tools that do more than run: they know and expose the limits within which they are right.

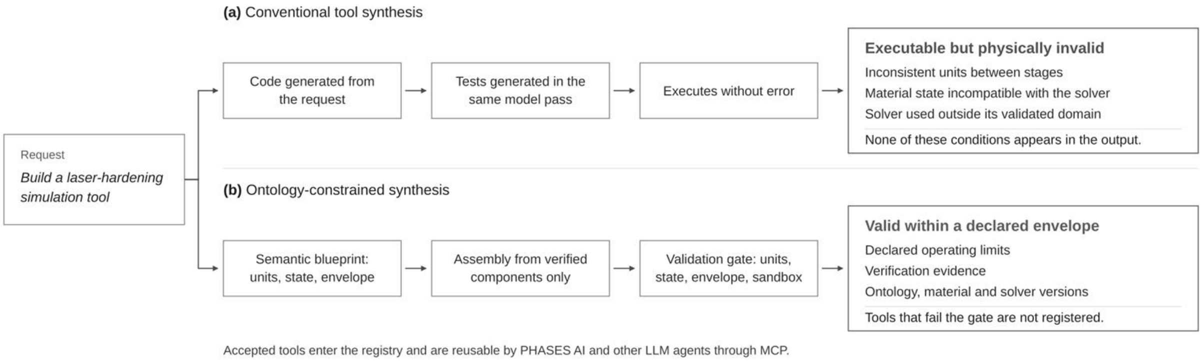

Figure 1. Conventional synthesis can produce a tool that runs yet remains physically invalid. PHASES uses an ontology-grounded blueprint, verified components, and a physics gate so accepted tools expose their operating limits and can be reused by PHASES AI and other LLM agents.

References

[1] Aerobase Innovations AB, “PHASES,” https://phases.aerobase.se (accessed 11 August 2026).
[2] European Materials Modelling Council, “Elementary Multiperspective Material Ontology (EMMO),” https://emmc.eu/emmo/ (accessed 11 August 2026).
[3] M. T. Horsch et al., “OSMO: Ontology for Simulation, Modelling, and Optimization,” CEUR Workshop Proceedings, vol. 2969, 2021, doi:10.5281/zenodo.5237774.
[4] T. Cai, X. Wang, T. Ma, X. Chen, and D. Zhou, “Large Language Models as Tool Makers,” International Conference on Learning Representations (ICLR), 2024.