

Can AI Agents Uncover Physics? Agentic Discovery of Atmospheric Relationships

Simon Driscoll (*Assistant Research Professor, DAMTP, University of Cambridge*) and Jon Crowcroft (*Professor, Department of Computer Science and Technology, University of Cambridge*)

Location: DAMTP/Computer Science

Primary contact: Dr Simon Driscoll (sd2136@cam.ac.uk)

Emails: sd2136@cam.ac.uk, jon.crowcroft@cl.cam.ac.uk

Abstract

AI agents - language models that search the literature, propose hypotheses, critique them and act on the results - have begun to produce scientific findings of their own: Google DeepMind's AlphaEvolve found new algorithms and improved bounds on open mathematical problems, and its Co-Scientist (Gottweis et al., 2026) proposed drug-repurposing candidates and new disease targets that were then validated in vitro. This project applies agentic AI to climate science, where AI models from Google DeepMind, Microsoft and NVIDIA now forecast the weather as well as the physics-based systems run by national weather centres.

A model can score well on standard forecast-error metrics without having learned the dynamics behind the data (Subel and Zanna, 2026). Physics-based models have conservation laws built in and are judged, beyond skill scores, on whether they reproduce the underlying physical processes. In contrast AI weather models have been judged almost entirely on error metrics, without such guarantees and the physical-consistency checks now being added (Kasteleyn et al., 2026) test relationships that a model satisfying the underlying physics would fulfill by construction.

In recent work, Lassen et al. (2026, arXiv) turned the question of whether a model reproduces the actual physical processes behind a phenomenon, which no conservation law guarantees, into a formal search problem: physicists define the physical quantities that matter, a search over decades of observations finds which of them, where and when, actually predict how the system evolves, and those relationships become fixed tests of whether an AI model has the dynamics right. Applied to the winter stratosphere, the search recovered the known precursors of a weakening polar vortex without being told them, and testing AI weather models against them showed which of these best preserved the underlying dynamics (not revealed by forecast scores alone).

This project asks whether an LLM agent can replace part of the pipeline in Lassen et al. (2026, arXiv). The pipeline already has the propose-and-verify structure that agentic discovery systems such as AlphaEvolve are built on, except that in Lassen et al. the proposal step is exhaustive enumeration over a space fixed in advance. An agent would propose candidates, learn from each result and decide what to try next. The first test is whether it can rediscover the physical drivers the exhaustive search found, which can be measured exactly, before extending the search beyond that space. The agent design, memory and language model are the student's to choose - the statistical guarantees must survive the switch to an agent.

Purpose-built agents for weather and climate so far analyse data, call forecast models as tools and occasionally run them; we are not aware of one that decides what physics a model must get right, or builds models for sound physics and dynamics. An agent that can take over this step could go on to extend the search from the literature and find new physical drivers in various observational datasets. It could also use theory from mathematics, physics and computer science (e.g. Boger et al., 2026, arXiv) to verify or reject models and potentially to help build more physically sound AI models. We envisage this project as an exciting first step on an open-ended route to agentic AI for scientific discovery in physics.

Further Details

There can be at most 1 student(s) on this project.

A working codebase and the data are available, with compute on CSD3 and A100 GPUs on JASMIN.

The project suits students interested in AI agents, large language models, machine learning, AI for scientific discovery, statistics and climate physics. The skills transfer directly to AI industry and academic research labs.

References

Evaluating Dynamical Fidelity through Predictive Structure in Physical Representations - <https://arxiv.org/abs/2609.34627> [[Oskar Bohn Lassen also available to contact to ask questions about the paper and help supervise if useful (obola@dtu.dk)]]

“Thinking Before Thinking: Scaling Agentic Inference Through Meta-Reasoning”
<https://arxiv.org/abs/2609.38147>

Accelerating scientific discovery with Co-Scientist_
<https://www.nature.com/articles/s41586-026-10644-y>

AlphaEvolve: A coding agent for scientific and algorithmic discovery -
<https://arxiv.org/abs/2506.13131>

Learning skillful medium-range global weather forecasting -
<https://www.science.org/doi/10.1126/science.adj2336>

PhysMetrics.Weather: An Evaluation Framework for Physical Consistency in ML Weather Models - <https://arxiv.org/abs/2606.10642>

A Response Theory Probe for Learned Stochastic AI Simulators, Tested on Lorenz-63 -
<https://arxiv.org/pdf/2610.06798>