Title: Teaching and Mentoring the AI Scientists

Speakers: Xiaoliang Qi

Collection/Series: Theory + AI Symposium

Date: April 08, 2025 - 1:40 PM

URL: https://pirsa.org/25040066

Abstract:

In the past two years, the LLM has made significant progress in math and reasoning, but it has not been applied widely in scientific research tasks. In this talk I will give a brief introduction to our on-going efforts on building the first AI scientist platform, where all researchers in different fields can contribute to teaching the AI scientists via contributing benchmarks and contributing specialized tools. We believe that by providing AI with the real-time updates of benchmarks and research tools, we are starting to enter an era with innovation driven by new types of human-AI collaboration.

Teaching and Mentoring AI Scientists

Xiao-Liang Qi

Path Integral Technology, Inc.

AI+Theory Conference at Perimeter Insitute

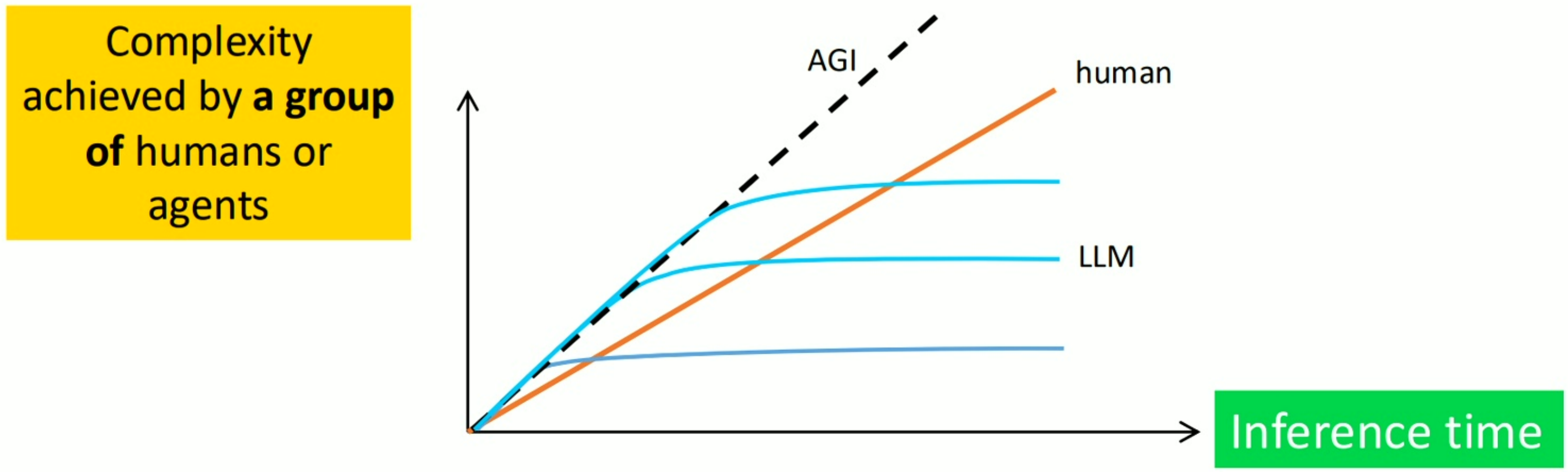
Apr 8th, 2025

Outline

- LLM agents in scientific research
- Our approach in teaching AI to do science:
  - Benchmarks
  - MCP servers
  - Scientific Research Agent: Lucien
- Further discussion

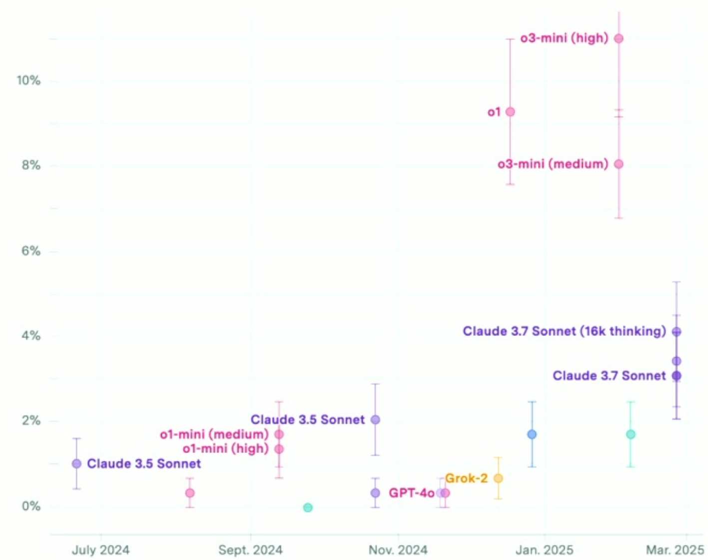
Complexity scaling law

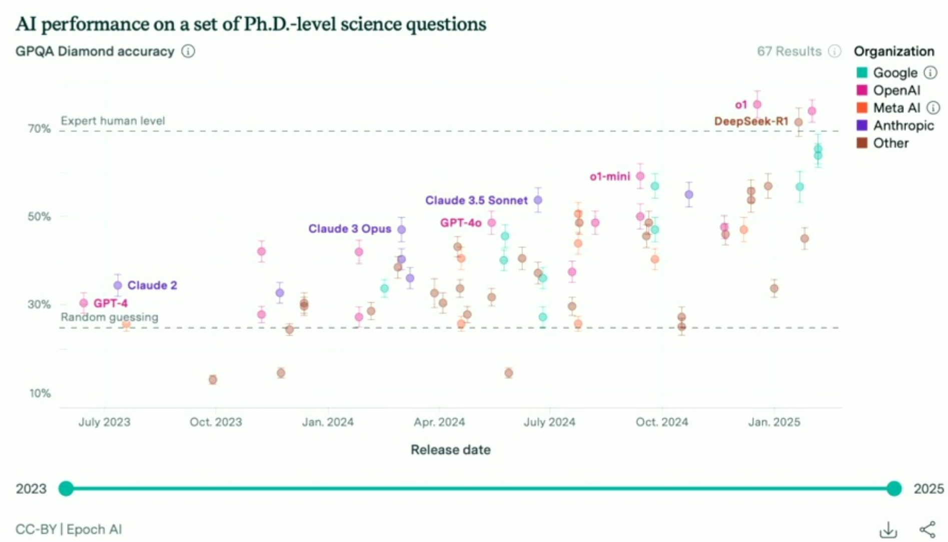
- The capability of an intelligent machine is measured by how complex is the task it can achieve
- Behind the “training scaling law” and “reasoning scaling law” is the **complexity scaling law**



LLM in Scientific Research

- Scientific innovation is a key milestone for human level AI.
- Two years "from primary school to graduate school"

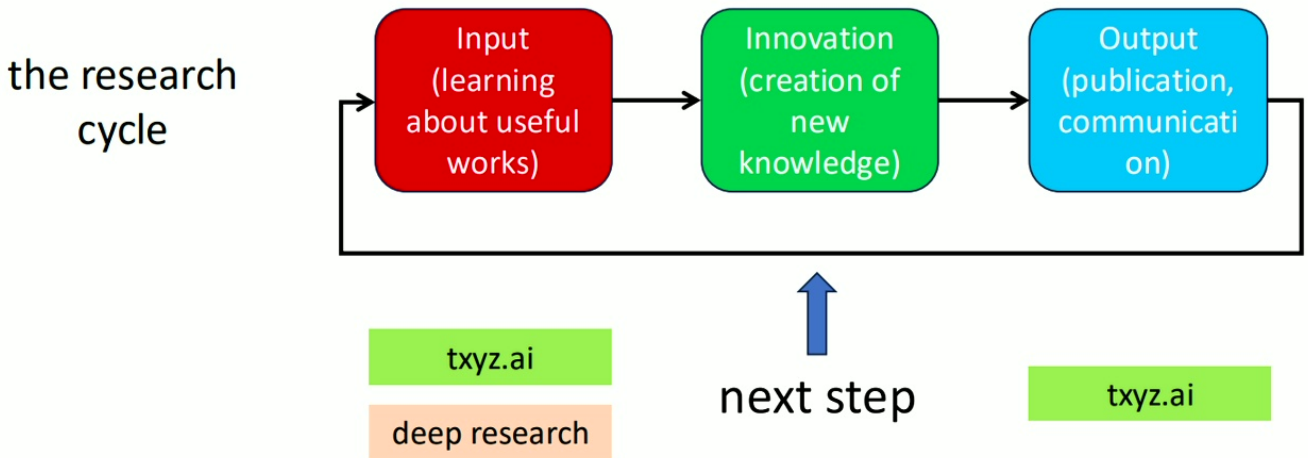




two example benchmarks from Epoch AI

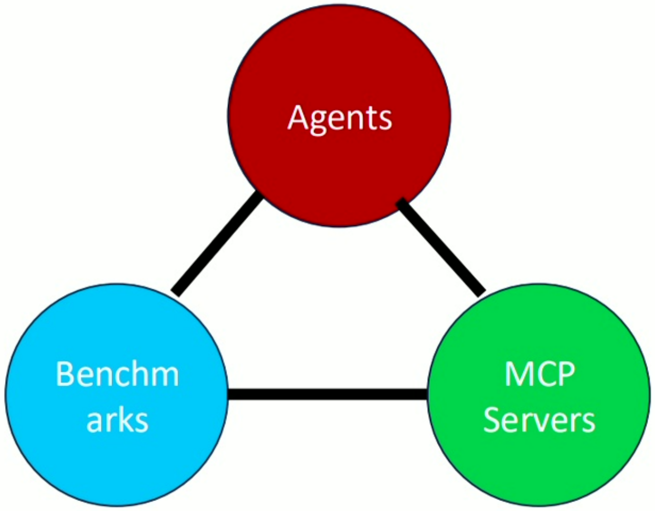
Towards the first AI Scientists Platform

- Our goal: build AI scientists at the human collaborator level.
- txyz.ai: AI native reading, search, writing tools. An arxivLabs project
- Building the AI scientist platform: Project Lucien

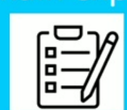


Our approach: build the AI with community

- Key capabilities of AI scientists:
  - Use of specialized tools
  - Integration with existing research workflow
  - Learn and evolve with the community (develop new tools, learn quickly from interaction with human)
- Building a platform that enables the entire research community to build benchmarks and tools



bench.science
Building benchmarks with domain experts



mcp.science
Building agent with domain-specific tools

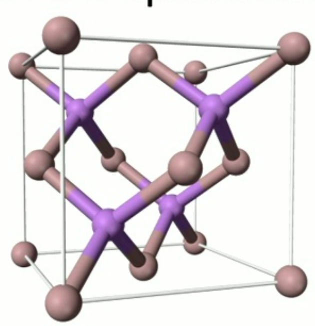
bench.science: The platform for building benchmarks for scientific research

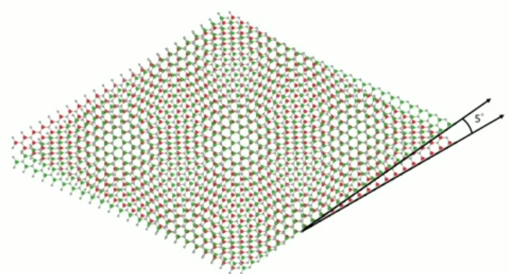
- Enables each expert to **seamlessly create an AI benchmark as a collaboration project**
- Inviting contributors and reviewers
- Automatically testing your questions on the SOTA LLM's
- Publishing the benchmark with a unique identifier
- Automatic evaluation of the benchmark on AI models
- Benchmarks are not public but verifiable
- Version management: Updating the benchmark set and the model evaluation results to reflect the current stage of research
- (Future plan) Contributions are tractable and rewarded

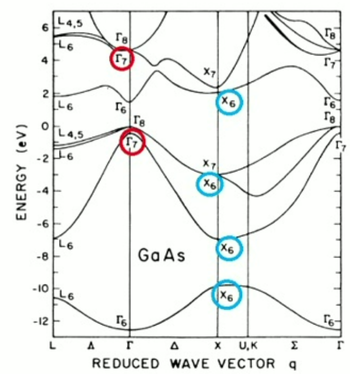
bench.science

Quantum Material Research Benchmark

- As a first example of research benchmarks, we collaborate with domain experts and develop **QMBench** in the field of quantum material computational research
- Diverse physical properties that are distinct for different materials
- Relatively mature numerical methods such as density functional theory
- A lot of computational tools developed

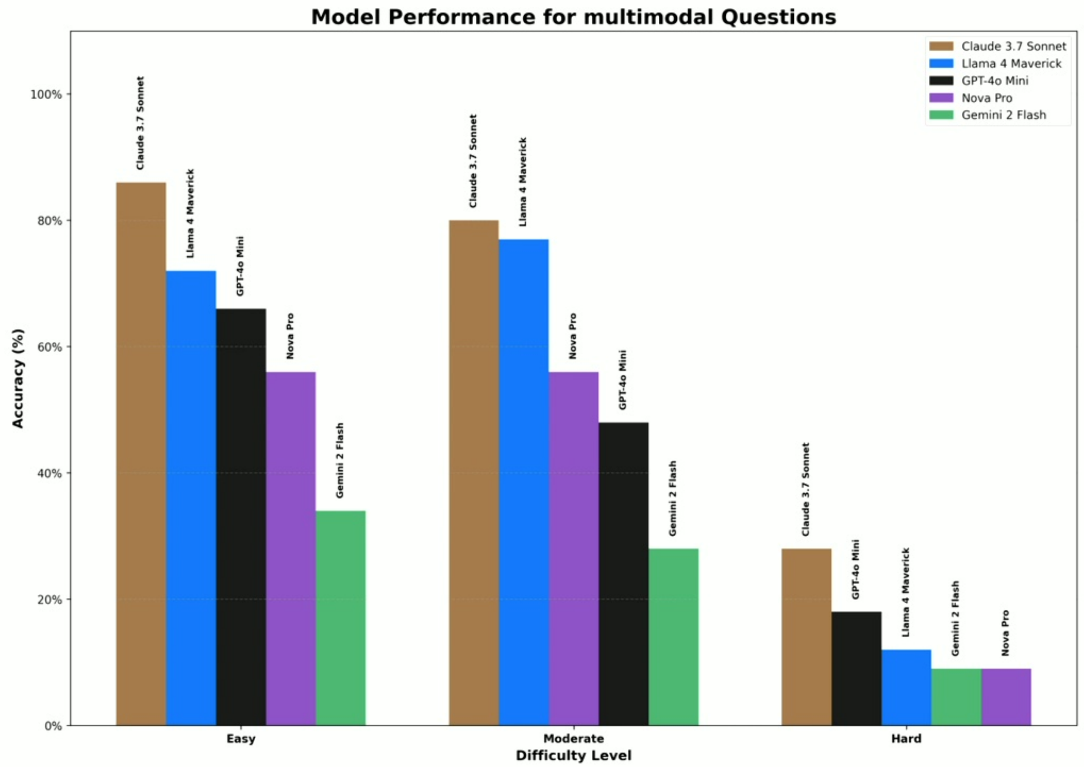






Yanzhen Wang, Yiyang Jiang, Chao-Xing Liu, XLQ, Binghai Yan,
https://bench.science/00A-202504-B

Performance of leading models in QMBench



https://bench.science

mcp.science: The platform for contributing research tools as MCP servers

- Model Context Protocol: an open-source protocol developed by Anthropic
- An "USB connector" between AI agent and tools
- This enables tool development to be independent from model development
- Example:
-

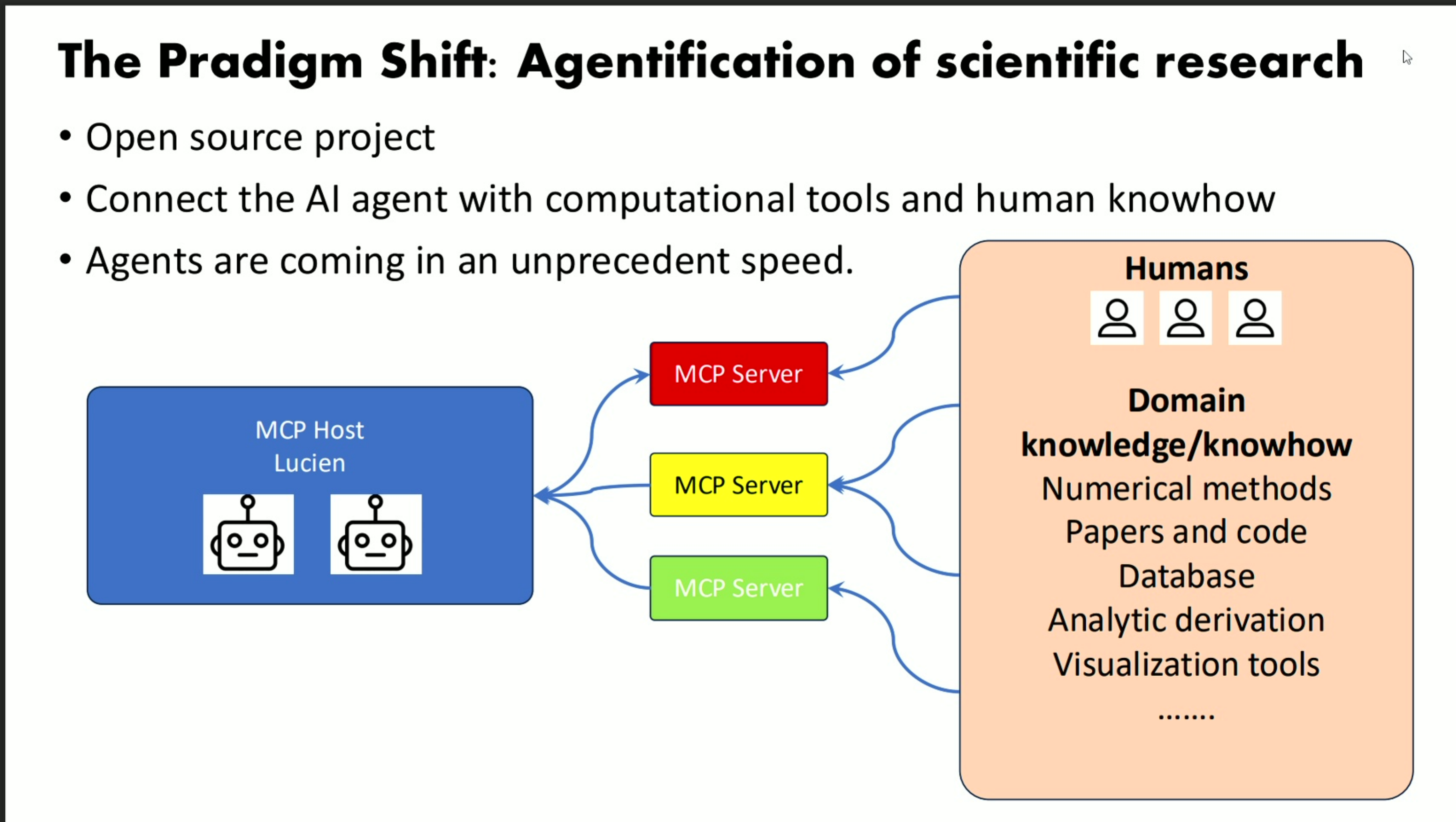
```
from mcp.server.fastmcp import FastMCP
from mcp.types import EmbeddedResource, ImageContent, TextContent, TextResourceContents
mcp: FastMCP = FastMCP(name="build_structure")

@mcp.tool()
async def search_materials_by_formula(chemical_formula: str) -> list[TextContent]:
    """
    Search for materials in the MPRester database by chemical formula

    Args:
        chemical_formula: the chemical formula of the material

    Returns:
        a list of TextContent objects, each TextContent object contains the
        description of a structure with the given chemical formula
    """

    search_results: Any = mpr.summary.search(formula=chemical_formula)
```



The Pradigm Shift: Agentification of scientific research
• Open source project
• Connect the AI agent with computational tools and human knowhow
• Agents are coming in an unprecedent speed.
MCP Host
Lucien
MCP Server
MCP Server
MCP Server
Humans
Domain
knowledge/knowhow
Numerical methods
Papers and code
Database
Analytic derivation
Visualization tools
....

Readme

Comprehensive Report on Narrow Band Semiconductors with Strong IR Absorption

Introduction

Narrow band semiconductors are crucial materials for infrared (IR) detection and emission applications. These materials possess small electronic bandgaps that correspond to photon energies in the infrared region, making them valuable for various applications including thermal imaging, night vision, spectroscopy, telecommunications, and environmental monitoring. This report provides a comprehensive analysis of key narrow bandgap materials with strong infrared absorption properties based on both literature review and first-principles calculations.

Key Narrow Bandgap Semiconductor Materials

1. Lead Chalcogenides (PbS, PbSe, PbTe)

Lead chalcogenides are an important class of narrow bandgap semiconductors with

Type your message here...

Built with Chainlit

An invitation for collaboration

- **Joining us in teaching the AI and collaborating with the AI**
- **Build benchmarks at** https://bench.science
- **Contribute your mcp server at** https://mcp.science
- **Sign up for testing Lucien for Quantum Material Research at** https://lucien.science
- bench.science and mcp.science are **open platforms** owned by the community.