# **Evolution of LLM-Based** Scientific Discovery

### **Ethics**

## **FUTURE**

Transparency

Self-Improvement

**Robotic Automation** 

Fully-Autonomous Cycle

#### **Machine Learning Research**



The AI Scientist (Lu et al., 2024)

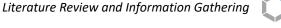


The AI Scientist-v2 (Yamada et al., 2025)



AI-Researcher (Data Intelligence Lab, 2025)

### LEGEND







**Experiment Planning and Execution** 



Conclusion and Hypothesis Validation

Strategic Iteration and Refinement

Data Analysis and Organization



Methodology / Framework



Benchmark / Evaluation



Agentic Workflow

### LEVEL 3 LLM as Scientist



Agent Laboratory (Schmidgall et al., 2025)

Zochi

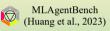
(Intology AI, 2025)



(Autoscience Institute, 2025)

**Data Modeling and Analysis** 

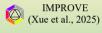
#### Machine Learning Research



BudgetMLAgent

(Gandhi et al., 2024)



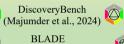


CodeScientist

(Jansen et al., 2025)











(Gu et al., 2024)

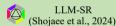


## Zheng et al., 2023

## LEVEL 2

LLM as Analyst

### **Function Discovery**





LLM-SRBench (Shojaee et al., 2025)



Gravity-Bench-v1 (Koblischke et al., 2025)



BoxLM (Li et al., 2024)

### Natural Science Research



(Ghafarollahi and Buehler, 2024) AI co-scientist



ScienceAgentBench (Chen et al., 2024)



Auto-Bench



(Chen et al., 2025) Gao et al., 2024



BioResearcher



DrugAgent (Liu et al., 2024) FutureHouse (Skarlinski et al., 2025)

(Gottweis et al., 2025)

Coscientist (Boiko et al., 2023)

### **General Research**



Liu et al., 2025



**CURIE** (Cui et al., 2025)

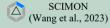


DiscoveryWorld (Jansen et al., 2024)



**EAIRA** (Cappello et al., 2025)

### **Literature Review & Info Gathering**





PaperQA (Lála et al., 2023)



LitLLM



(Agarwal et al., 2024)



Dennstädt et al., 2024





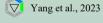
ARCADE

(Li et al., 2025)

(Jiang et al., 2024) arXiv2Table

(Wang et al., 2025)

(Luo et al., 2024)



IdeaBench

GraphEval

(Feng et al., 2025)

AI Idea Bench 2025

ResearchBench

(Liu et al., 2025)



(Yang et al., 2024) SciAgents

Nova

(Hu et al., 2024)

Buehler, 2024



(Ghafarollahi et al., 2024)

MOOSE-Chem



KG-CoI

**Idea Generation & Hypothesis Formulation** 



Qi et al., 2024

∇ Oi et al., 2023



Scideator (Radensky et al., 2024)

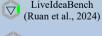
△ Ciuca et al., 2023



LiveIdeaBench



FieldSHIFT (O'Brien et al., 2024)



Zhou et al., 2024



LLM4GRN (Afonja et al., 2024)

**Iteration & Refinement** 

#### **Experiment Planning & Execution**

Science Hierarchography

(Gao et al., 2025)

(Newman et al., 2024)



LEVEL 1

LLM as Tool

DS-1000 (Lai et al., 2022)

BioPlanner

(O'Donoghue et al., 2023)



CharXiv

(Wang et al., 2024)

TableBench

(Wu et al., 2024)



**Data Analysis & Organization** 

AutomaTikZ. (Masry et al., 2022) (Belouadi et al., 2023)



ChartX & ChartVLM (Xia et al., 2024)



(Zadeh et al., 2024)

### **Conclusion & Hypothesis Validation**



ReviewerGPT (Liu et al., 2023) ReviewCritique (Du et al., 2024)



Takagi et al., 2023

Tyser et al., 2024



RR-MCQ (Zhou et al., 2024)



Xu et al., 2025 ClaimCheck (Ou et al., 2025)



MC-NEST (Rabby et al., 2025)

Explanation-Refiner

(Quan et al., 2024)

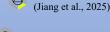
Chain-of-Ideas

(Li et al., 2024)



MLE-Bench

(Chan et al., 2024)



Shi et al., 2025







(Wang et al., 2024)

## Chain-of-Table

SciReplicate-Bench (Xiang et al., 2025)

#### CycleResearcher (Weng et al., 2024)