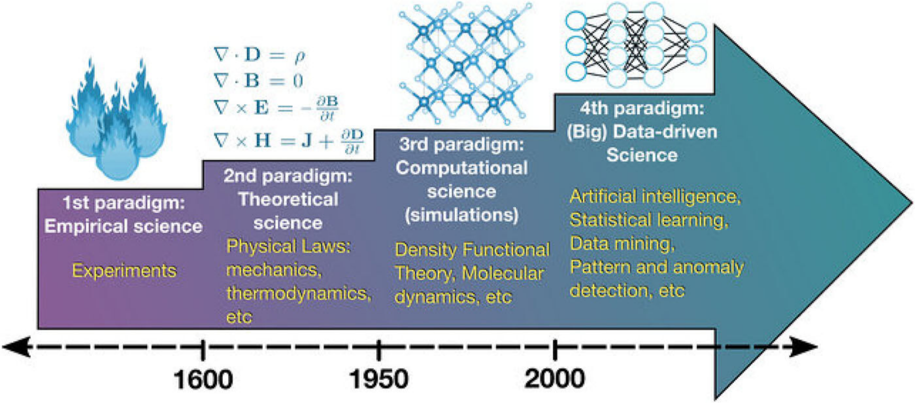
From Prediction to Discovery

Agentic AI for Science

Armando Vieira

May 27, 2026

Science evolution



Science done differently

Article | Published: 19 May 2026

Accelerating scientific discovery with Co-Scientist

Juraj Gottweis ✉, Wei-Hung Weng ✉, Alexander Daryin, Tao Tu, Petar Sirkovic, Artiom Myaskovsky, Grzegorz Glowaty, Felix Weissenberger, Alessio Orlandi, Dan Popovici, Anil Palepu, Keran Rong, Ryutaro Tanno, Khaled Saab, Fan Zhang, Jacob Blum, Andrew Carroll, Kavita Kulkarni, Nenad Tomašev, Dina Zverinski, Ivor Rendulic, Elahe Vedadi, Florian Hasler, Luka Rimanic, … Vivek Natarajan ✉

Article | Published: 19 May 2026

A multi-agent system for automating scientific discovery

Ali Essam Ghareeb, Benjamin Chang, Ludovico Mitchener, Angela Yiu, Caralyn J. Szostkiewicz, Dmytro Shved, Gavin J. Gyimesi, Jon M. Laurent, Samantha M. Wright, Muhammed T. Razzak, Andrew D. White ✉, Silvia C. Finnemann, Michaela M. Hinks ✉ & Samuel G. Rodriques ✉

Article | Published: 19 May 2026

An AI system to help scientists write expert-level empirical software

Eser Aygün, Anastasiya Belyaeva, Gheorghe Comanici, Marc Coram, Hao Cui, Jake Garrison, Renee Johnston, Anton Kast, Cory Y. McLean, Peter Norgaard, Zahra Shamsi, David Smalling, James Thompson, Subhashini Venugopalan, Brian P. Williams, Chujun He, Sarah Martinson, Martyna Plomecka,

AI changes the structure of exploration

AI is not just a faster calculator.

It changes the geometry of scientific search.

AI systems can help us:

- traverse enormous conceptual spaces;
- generate hypotheses;
- connect distant domains;
- simulate possibilities;
- iterate much faster than individual human cognition.

Science becomes less like a pipeline
and more like a hybrid cognitive ecosystem.

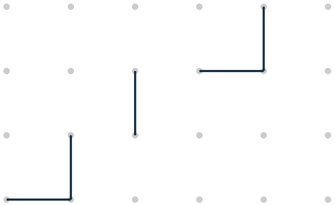
Core idea

AI is not merely a new tool for science.

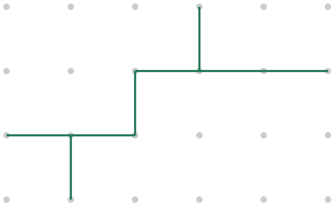
It may reveal that science itself was always
a form of guided, collective, coherence-seeking exploration.

Not just a method for knowing the world, but a way the world generates new forms of knowing.

Science as Percolation



Early: isolated ideas



Threshold: connected framework

In percolation, local links suddenly become a global path.

In science, isolated results can suddenly connect into a new framework.

Agentic AI can speed up and improve drastically the range of these connections: literature, simulations, hypotheses, and experiments.

Incremental Science vs. Disruptive Science

Most science is incremental

- Better measurements
- Cleaner datasets
- Smaller error bars
- Improved controls
- Better simulations
- Slightly better molecules or materials

Rarely, science is disruptive

- A new framework changes the map
- Old facts are reorganized
- New questions become possible
- Example 1: Quantum Mechanics
- Example 2: Theory of Evolution of Species

Breakthroughs are rare. Most breakthroughs stand on thousands of small steps.

Traditional Science vs. Agentic Science

Old workflow:

"I have an idea. AI, please help me check it."

New workflow:

"AI, generate strange ideas, attack my assumptions, find contradictions, and tell me why my favorite hypothesis is probably wrong."

Key shift: the scientist is no longer only the commander.
The scientist becomes a **chaotic promoter of useful collisions**.

The Scientist Is Not the Orchestra Conductor Anymore

Traditional model:

Scientist $\rightarrow$ Tool $\rightarrow$ Result

Agentic model:

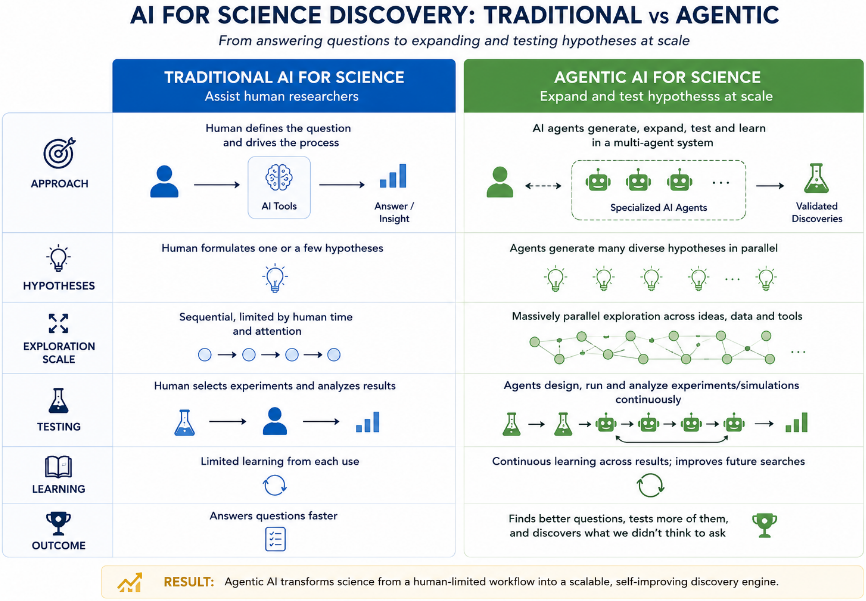
Scientist $\leftrightarrow$ Agents $\leftrightarrow$ Data, Literature, Simulations, Experiments

A less elegant but more accurate image

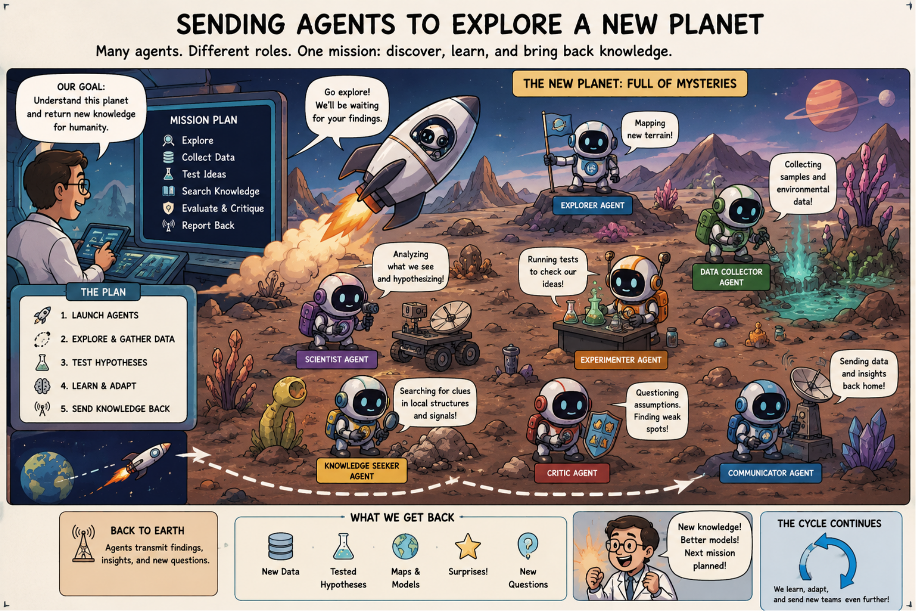
The scientist is now someone releasing intelligent raccoons into a library and saying:

"Find me something interesting, but please do not burn the building."

Agentic vs Non-agentic AI



Agentic vs Non-agentic AI



Traditional AI vs Agentic AI for Scientific Discovery

Traditional AI

Goal: Assist the researcher

Typical Prompt

> *"Summarize recent papers on protein aggregation and suggest 3 hypotheses."*

Characteristics

- Human drives the workflow
- AI answers specific questions
- Limited exploration
- Sequential reasoning
- Few hypotheses tested

Workflow

Researcher → AI Tool → Insight

Agentic AI

Goal: Expand and test hypotheses at scale

Typical Prompt

> *"Generate competing hypotheses, test them with simulations and literature review, critique weak ideas, refine promising ones, and iterate autonomously."*

Characteristics

- Multi-agent collaboration
- Autonomous exploration
- Parallel hypothesis generation
- Continuous testing and refinement
- Self-critique and adaptation

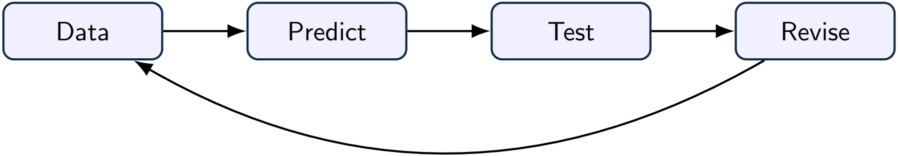
From Prediction to Discovery

Predictive AI

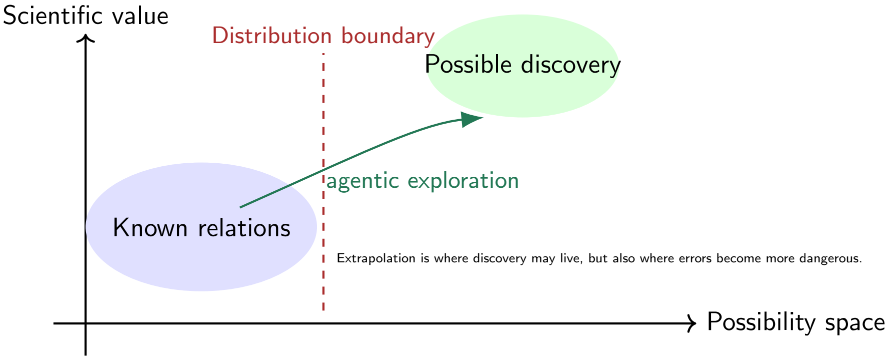
- Learns from existing data
- Predicts likely outcomes
- Works best inside known distributions
- Answers: "What is probable?"

Agentic discovery

- Plans multi-step workflows
- Uses scientific tools
- Tests hypotheses
- Explores possibility space
- Asks: "What should we try next?"



Out-of-Distribution: Where Discovery Often Lives



- New drugs may not look like previous drugs.
- New materials may sit far from known examples.
- New mechanisms may not fit old categories.

The challenge: explore beyond the training distribution without confusing novelty with nonsense.

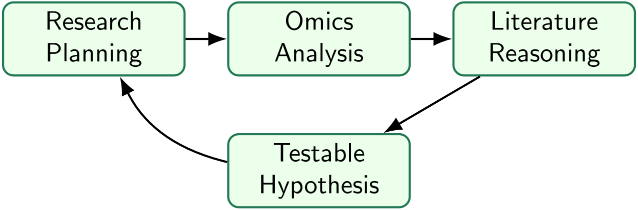
Where Agentic AI Can Improve Discovery

- **Literature synthesis:** map papers into mechanisms, conflicts, and gaps.
- **Hypothesis generation:** propose causal explanations and testable predictions.
- **Experiment design:** choose variables, controls, and next-best experiments.
- **Simulation orchestration:** run molecular dynamics, DFT, plasma, quantum, or climate simulations.
- **Data analysis:** clean, analyze, debug, and interpret complex datasets.
- **Closed-loop discovery:** connect prediction, experiment, feedback, and model updating.
- **Cross-domain reasoning:** connect biology, chemistry, physics, and engineering.

Biology Example: Therapeutic Discovery Agents

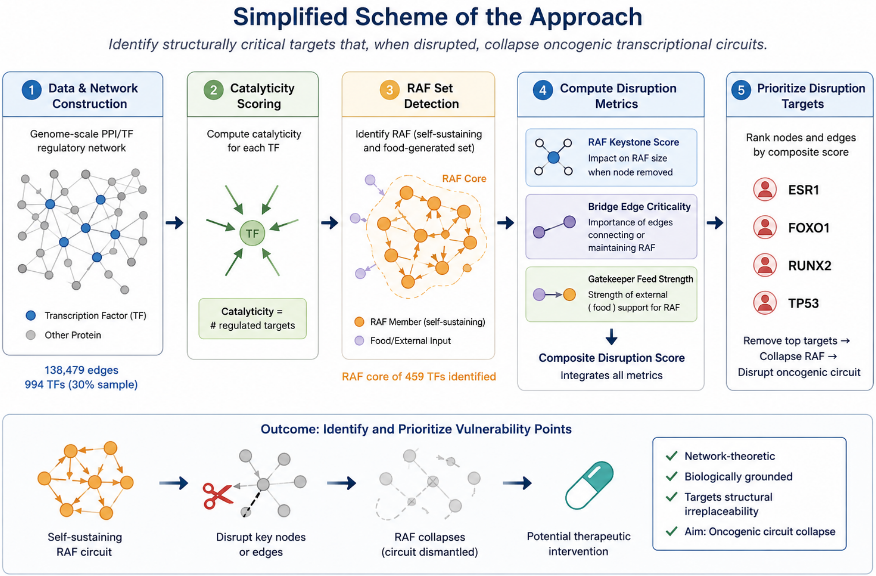
Agentic AI can support:

- Target identification from multi-omics data
- Disease mechanism reasoning
- Drug response prediction
- Literature-grounded hypothesis generation
- Candidate prioritization before wet-lab validation

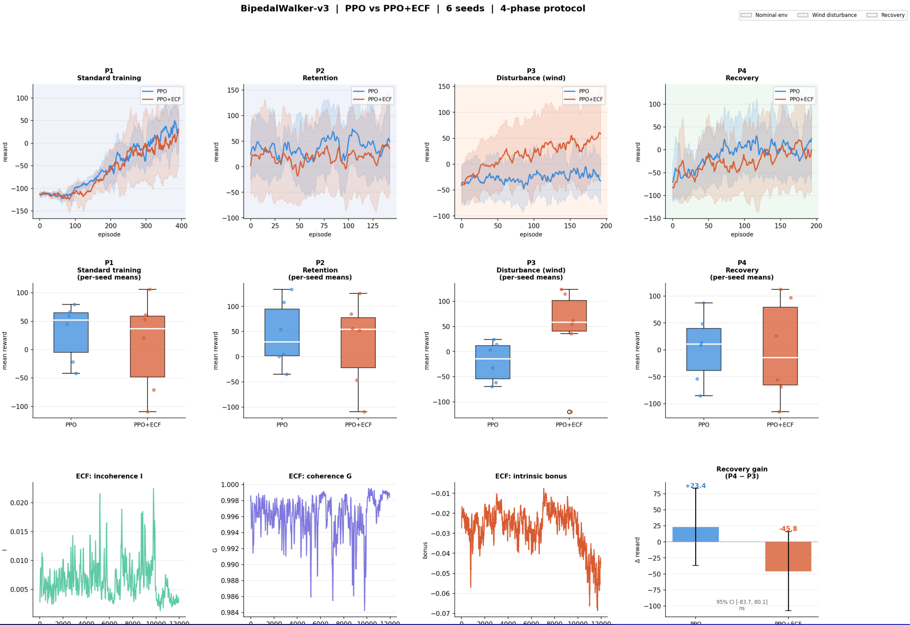


Source / inspiration: MEDEA and PandaClaw agentic therapeutic discovery systems.

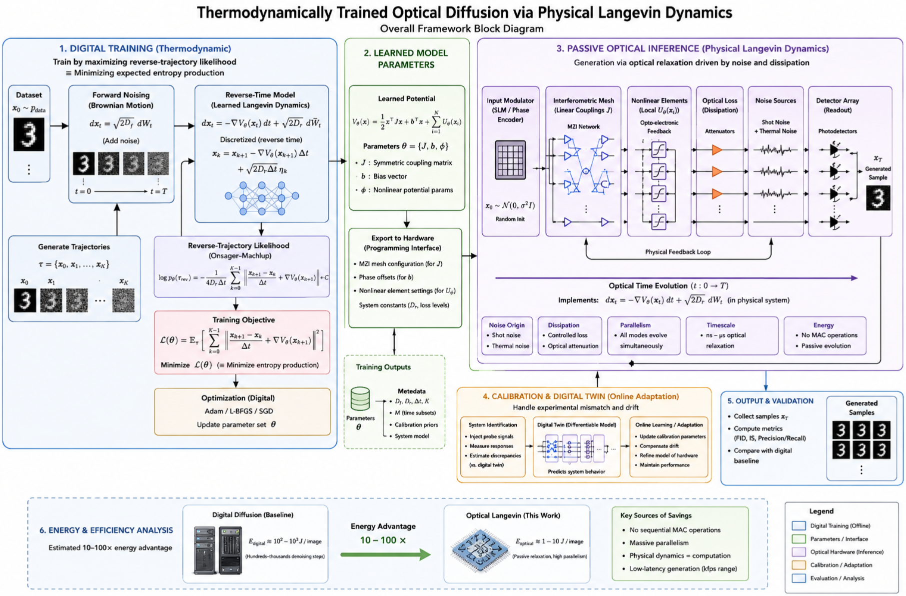
Success case: GRN in Cancer



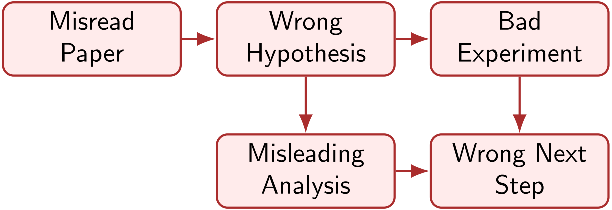
Success case: Reinforcement learning with curiosity



Failure Case: Optical diffusion



Cascading Scientific Errors



One small error becomes an automated family tree of wrong conclusions.

Physics analogy: resonance. A small vibration, repeated at the wrong frequency, can break the bridge.

Science is not one query. It is a distributed process.

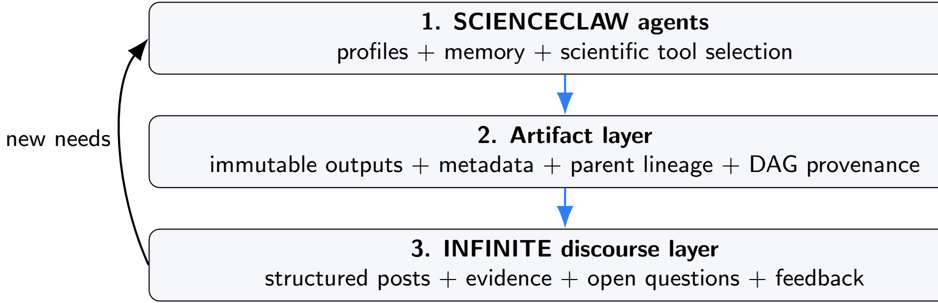
A shared ecosystem where independent agents:

- select scientific tools according to their profile;
- produce immutable computational artifacts;
- broadcast unresolved information needs;
- synthesize compatible results from other agents;
- publish traceable findings in a structured scientific discourse layer.

Takeaway

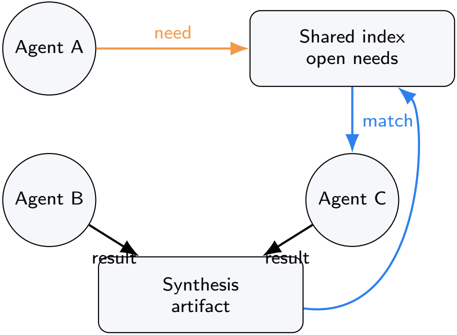
The acceleration comes less from a single brilliant model and more from **parallel, traceable, tool-using scientific loops**.

System architecture: three layers



This is closer to a scientific laboratory ecosystem than a chatbot: tools, records, partial results, criticism, and iteration.

ArtifactReactor: coordination without a central planner

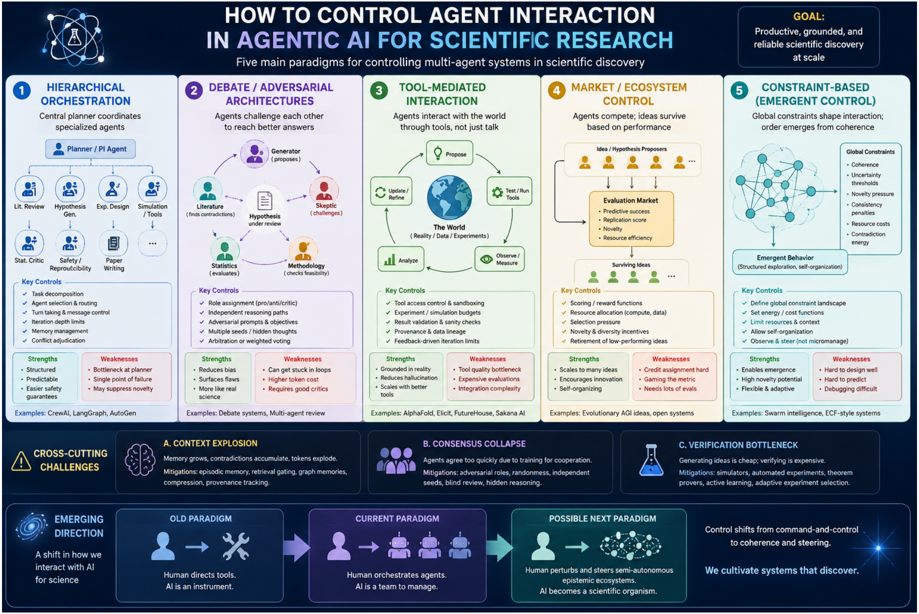


Mechanism

1. One agent hits a gap.
2. The gap becomes a public need.
3. Other agents detect matches.
4. New artifacts are generated.
5. Compatible artifacts are fused.

Why this matters: idle hypotheses become actionable requests instead of disappearing in a chat log.

control



HOW TO CONTROL AGENT INTERACTION
IN AGENTIC AI FOR SCIENTIFIC RESEARCH
Five main paradigms for controlling multi-agent systems in scientific discovery
GOAL:
Productive, grounded, and reliable scientific discovery at scale
1 HIERARCHICAL ORCHESTRATION
Central planner coordinates specialized agents
Planner / PI Agent
Lit. Review
Hypothesis Gen.
Exp. Design
Simulation / Tools
Stat. Critic
Safety / Reproducibility
Paper Writing
...
Key Controls
Task decomposition
Agent selection & routing
Turn taking & message control
Iteration depth limits
Memory management
Conflict adjudication
Strengths
Structured
Predictable
Easier safety guarantees
Weaknesses
Bottleneck at planner
Single point of failure
May suppress novelty
Examples: CrewAI, LangGraph, AutoGen
2 DEBATE / ADVERSARIAL ARCHITECTURES
Agents challenge each other to reach better answers
Generator (proposes)
Literature (finds contradictions)
Hypothesis under review
Skeptic (challenges)
Statistics (evaluates)
Methodology (checks feasibility)
Key Controls
Role assignment (pro/anti/critic)
Independent reasoning paths
Adversarial prompts & objectives
Multiple seeds / hidden thoughts
Arbitration or weighted voting
Strengths
Reduces bias
Surfaces flaws
More like real science
Weaknesses
Can get stuck in loops
Higher token cost
Requires good critics
Examples: Debate systems, Multi-agent review
3 TOOL-MEDIATED INTERACTION
Agents interact with the world through tools, not just talk
Propose
Update / Refine
Test / Run Tools
The World (Reality / Data / Experiments)
Analyze
Observe / Measure
Key Controls
Tool access control & sandboxing
Experiment / simulation budgets
Result validation & sanity checks
Provenance & data lineage
Feedback-driven iteration limits
Strengths
Grounded in reality
Reduces hallucination
Scales with better tools
Weaknesses
Tool quality bottleneck
Expensive evaluations
Integration complexity
Examples: AlphaFold, Elicit, FutureHouse, Sakana AI
4 MARKET / ECOSYSTEM CONTROL
Agents compete; ideas survive based on performance
Idea / Hypothesis Proposers
Evaluation Market
Predictive success
Replication score
Novelty
Resource efficiency
Surviving Ideas
Key Controls
Scoring / reward functions
Resource allocation (compute, data)
Selection pressure
Novelty & diversity incentives
Retirement of low-performing ideas
Strengths
Scales to many ideas
Encourages innovation
Self-organizing
Weaknesses
Credit assignment hard
Gaming the metric
Needs lots of evals
Examples: Evolutionary AGI ideas, open systems
5 CONSTRAINT-BASED (EMERGENT CONTROL)
Global constraints shape interaction; order emerges from coherence
Global Constraints
Coherence
Uncertainty thresholds
Novelty pressure
Consistency penalties
Resource costs
Contradiction energy
Emergent Behavior (Structured exploration, self-organization)
Key Controls
Define global constraint landscape
Set energy / cost functions
Limit resources & context
Allow self-organization
Observe & steer (not micromanage)
Strengths
Enables emergence
High novelty potential
Flexible & adaptive
Weaknesses
Hard to design well
Hard to predict
Debugging difficult
Examples: Swarm intelligence, ECF-style systems
CROSS-CUTTING CHALLENGES
A. CONTEXT EXPLOSION
Memory grows, contradictions accumulate, tokens explode.
Mitigations: episodic memory, retrieval gating, graph memories, compression, provenance tracking.
B. CONSENSUS COLLAPSE
Agents agree too quickly due to training for cooperation.
Mitigations: adversarial roles, randomness, independent seeds, blind review, hidden reasoning.
C. VERIFICATION BOTTLENECK
Generating ideas is cheap; verifying is expensive.
Mitigations: simulators, automated experiments, theorem provers, active learning, adaptive experiment selection.
EMERGING DIRECTION
A shift in how we interact with AI for science
OLD PARADIGM
Human directs tools.
AI is an instrument.
CURRENT PARADIGM
Human orchestrates agents.
AI is a team to manage.
POSSIBLE NEXT PARADIGM
Human perturbs and steers semi-autonomous epistemic ecosystems.
AI becomes a scientific organism.
Control shifts from command-and-control to coherence and steering.
We cultivate systems that discover.

How it can speed up science

Speed mechanism	Effect on scientific work
Parallel agents	Many independent lines of reasoning run at once.
Tool chaining	Literature, simulation, databases, statistics, and visualization can be composed automatically.
Artifact reuse	Previous outputs become searchable building blocks for later investigations.
Open-needs broadcast	A failed or incomplete step becomes a task for another specialist agent.
Provenance DAG	Faster checking: humans can inspect how a claim was produced.
Structured discourse	Findings include evidence, uncertainty, and follow-up questions.

Main acceleration principle

Science gets faster when the loop **hypothesis** → **test** → **artifact** → **critique** → **next test** becomes continuous and machine-readable.

Why this is powerful, but not magic

What it improves

- exploration speed;
- literature and database coverage;
- reproducibility of computational steps;
- memory across investigation cycles;
- human auditability.

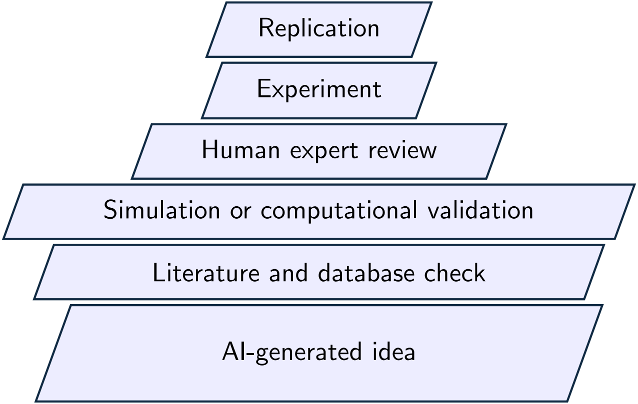
What it does not solve alone

- experimental validation;
- false assumptions in tools or data;
- overproduction of weak hypotheses;
- lack of strong baselines;
- confusing activity with discovery.

Critical view

The platform may greatly accelerate **iteration**, but truth still requires validation, comparison, and independent replication.

Guardrails: Hypothesis Testing Before Celebration



Rule: an AI-generated hypothesis is not a discovery until it survives testing.

What Good Scientific Agents Should Report

A useful scientific agent should not only answer.

It should report:

- What hypothesis it generated
- What evidence supports it
- What evidence contradicts it
- What assumptions it made
- How uncertain it is
- Whether the result is in-distribution or out-of-distribution
- What experiment could falsify it

Good science is not confidence. Good science is disciplined doubt.

Practical Roadmap for Scientific Agentic AI

Phase	Use	Guardrail
0	Literature search and summarization	Source and citation validation
1	Hypothesis generation	Human expert review
2	Simulation planning and analysis	Reproducible logs and parameter tracking
3	Closed-loop experiment suggestions	Approval before physical experiments
4	Autonomous lab integration	Safety, audit, replication, and governance

Autonomy should be earned, not assumed.

Summary

- Science is a loop: question, hypothesis, test, error, revision.
- Most science is incremental, but rare disruptions can reorganize the map.
- AI is moving from **prediction** toward **agentic discovery**.
- Agentic AI can search vast possibility spaces in biology, chemistry, physics, and materials science.
- Out-of-distribution exploration is where discoveries may live, but also where errors become more dangerous.
- The biggest risk is confusing plausible outputs with validated knowledge.

Let AI explore possibilities.
Let science decide what is real.

Selected Sources and Inspirations

Insilico Medicine. *Insilico Medicine Launches PandaClaw: Empowering Biologists with Agentic AI for Therapeutic Discovery*. 2026.
https://insilico.com/news/spjz8fzmb1-insilico-medicine-launches-pandaclaw-emp

MIMS Harvard. *MEDEA: An omics AI agent for therapeutic discovery*. GitHub repository.
https://github.com/mims-harvard/MEDEA

Szymanski et al. *An autonomous laboratory for the accelerated synthesis of novel materials*. Nature, 2023.
https://www.nature.com/articles/s41586-023-06734-w

Google DeepMind. *Millions of new materials discovered with deep learning*. 2023.
https://deepmind.google/discover/blog/millions-of-new-materials-discovered-with-deep-learning/

AUTONOMOUS AGENTS COORDINATING DISTRIBUTED DISCOVERY THROUGH EMERGENT ARTIFACT EXCHANGE .
https://arxiv.org/pdf/2603.14312

Zechen Zhang. X thread on scientific workflows and agentic research ideas.
https://x.com/ZechenZhang5/status/2049991768167649483