BRIEFING 02 · AI & DISCOVERY

Autonomous hypothesis, research and discovery — from DARPA to HAYAT

A deep dive into AI that does science: from the first robot scientist to the 2026 'AI scientist' papers, DARPA's expMath and MATHBAC, AlphaEvolve, and how I build verifiable research into HAYAT.

Mohsin Khan Shaik
AI Infrastructure & Site Reliability Engineer · San Francisco · mohsinkhanshaik.com
Edition of 2026-11-22

02 · AI & DISCOVERY

Autonomous hypothesis, research and discovery — from DARPA to HAYAT

A deep dive into AI that does science: from the first robot scientist to the 2026 'AI scientist' papers, DARPA's expMath and MATHBAC, AlphaEvolve, and how I build verifiable research into HAYAT.

2026	48	expMath	7,865
AI-scientist papers in Nature	multiplications: AlphaEvolve's 4×4 complex result	DARPA's AI co-author for math	tests behind HAYAT's agent loop

Key takeaways

- AI now takes part in every step of research: reading literature, proposing hypotheses, planning experiments, and writing papers.
- 2026 was a milestone year: Nature published papers on Google's AI co-scientist, FutureHouse's Robin and Sakana's AI Scientist, while critics note humans still did the lab work and much output lacks novelty.
- DARPA is funding AI co-authors for mathematics (expMath) and self-improving agentic discovery platforms (MATHBAC).
- The bottleneck is shifting from generating ideas to verifying them: provenance, citations and honest uncertainty.
- HAYAT applies the same principles locally: cited research, background learning, action verification and 7,865 automated tests.

A short history of machines doing science

- **1970s–80s:** expert systems such as DENDRAL inferred molecular structures from mass-spectrometry data.
- **2009:** the robot scientist "Adam" (Aberystwyth University) autonomously generated and tested hypotheses about yeast genes.
- **2020–2024:** AlphaFold predicted protein structures at near-experimental accuracy, recognized by the 2024 Nobel Prize in Chemistry.
- **2025:** DeepMind's AlphaEvolve, an LLM-driven evolutionary coding agent, found a way to multiply 4×4 complex-valued matrices with 48 scalar multiplications, improving on Strassen's 1969 approach for that case, and sped up a Gemini kernel by 23% [6].
- **2026:** "AI scientist" systems are published in Nature [1].

How an AI scientist works

Most systems are **multi-agent loops** built on large language models:

1. **Literature agent:** searches and summarizes papers, extracting claims and evidence.

2. **Hypothesis agent:** proposes ideas, often many in parallel, and ranks them through debate or tournament-style comparison.
3. **Planning agent:** designs experiments or analyses; in self-driving labs this becomes robot instructions.
4. **Analysis agent:** runs code on data, checks statistics, and updates beliefs.
5. **Writing and review agents:** draft papers and critique them before humans see them.

What separates useful systems from paper generators is **verification**: grounding every claim in a source, running real code or experiments, and measuring how often predictions turn out right.

State of the art in 2026

- **Google AI co-scientist:** built on Gemini, with validations reported for acute myeloid leukaemia drug candidates and liver-fibrosis targets [1].
- **FutureHouse Robin:** a multi-agent system whose best-known result is proposing ripasudil, a glaucoma drug, as a candidate for dry age-related macular degeneration; wet-lab work was done by humans [1].
- **Sakana AI Scientist:** produced a paper that scored above the average human acceptance threshold at a workshop, a lower bar than a main conference track [1].
- **Criticism:** commentators argue much AI-generated research is of low quality and novelty, and an early study found chatbots generated few insightful hypotheses in vaccinology [1].

What DARPA is funding

- **expMath (Exponentiating Mathematics):** aims to radically accelerate pure mathematics with an AI co-author that proposes and proves useful abstractions. It has teams on AI methods and on evaluation against professional-level problems; it kicked off in 2025 [2][3].
- **GENIUS (UMD + MIT, 2026):** a US$2.6M expMath effort building a neuro-symbolic reasoning engine that embeds mathematical strategies into language models [4].
- **MATHBAC (2026):** targets next-generation agentic AI platforms that systematically improve themselves to transform scientific discovery for defense [5].
- **DICE (2026):** studies how up to 100,000 AI agents can coordinate without a central orchestrator, the "society of agents" behind large-scale discovery [7].
- Earlier, **ASKE** and **ASKEM** worked on automatically extracting and modeling scientific knowledge, and the **AI Cyber Challenge** showed AI systems finding and fixing software vulnerabilities.

How HAYAT approaches autonomous research

HAYAT, my open-source, local-first assistant, is a practical, small-scale version of these ideas running on one PC [8]:

- **Research with receipts:** background web research cites its sources and checks freshness and conflicts; web content is always treated as untrusted data.
- **Learning gaps:** topics HAYAT couldn't answer can be researched in the background and remembered (with a switch to turn it off).
- **Predictions with track records:** HAYAT keeps measured track records for its predictions instead of just sounding confident.
- **Proof of action:** it only says something was done when a tool result shows it, and outward-facing actions need approval.

- **Verified engineering:** 597 modules and 7,865 automated tests keep the agent loop, memory and tools honest.

The design rule I took from building it: an autonomous system earns trust by showing its work, not by sounding smart.

How it could change life

- **Health:** faster drug repurposing, where an AI finds an approved drug that might treat another disease, can bring treatments to patients years sooner.
- **Materials and energy:** AI-guided discovery of battery chemistries, catalysts and alloys.
- **Mathematics and software:** AI co-authors that suggest lemmas, find counterexamples and optimize algorithms (as AlphaEvolve did).
- **Access:** small teams, startups and universities with limited budgets can run research programs that once needed a large lab.

Future scenarios

2027	2030	2035	2040+
AI co-scientists are standard in pharma and materials R&D, with humans choosing what to test.	Closed-loop self-driving labs in narrow domains: AI proposes, robots test, AI updates.	AI co-authorship is routine in math and science; verification capacity is the bottleneck.	Discovery at "agent-society" scale, with strong provenance standards for machine-generated science.

- **Optimistic:** discovery rates rise sharply in medicine and materials while review systems keep quality high.
- **Base case:** AI dramatically speeds the routine parts of research; breakthroughs still need human insight.
- **Pessimistic:** floods of low-quality AI papers overwhelm peer review and erode trust.

Effects and risks

- **Quality and trust:** cheap papers can drown good ones; journals need provenance rules and AI-use disclosure.
- **Reproducibility:** every AI-generated claim should come with code, data and logs.
- **Dual use:** automated discovery in biology and chemistry needs safeguards and access controls.
- **Credit:** who gets authorship when an AI proposed the idea?

What to watch

- Independent replication of AI-proposed findings in wet labs.
- expMath results on professional-level math problems.
- Journal policies on AI-generated research and reviewer workload.
- Open benchmarks for hypothesis quality, not just paper fluency.

My take

The next decade of AI for science will be won on verification. The systems I trust most will cite every claim, run real experiments, log everything, and say "I don't know" when that's the truth. It's the same

standard I hold HAYAT to, and the same standard an SRE holds a production system to: if you can't observe it and prove it, it didn't happen.

Key players & programs

Google DeepMind	AI co-scientist (Nature, 2026); AlphaEvolve (2025); AlphaFold [1][6]
FutureHouse / Edison Scientific	Robin multi-agent discovery system (Nature, 2026) [1]
Sakana AI	The AI Scientist, end-to-end automated papers (Nature, 2026) [1]
DARPA	expMath, MATHBAC, DICE, and earlier ASKE/ASKEM [2][5][7]
HAYAT (open source)	Local-first assistant with cited research, background learning and verified actions [8]

Frequently asked questions

What is an AI scientist?
A system, usually several AI agents working together, that reads literature, proposes hypotheses, designs experiments or analyses, and writes up results, with humans supervising key steps.

Has AI made real scientific discoveries?
Yes, with caveats. AlphaFold transformed protein-structure prediction, AlphaEvolve improved a matrix-multiplication algorithm, and 2026 AI-scientist papers reported drug-repurposing candidates, though humans did the lab validation.

What is DARPA expMath?
Exponentiating Mathematics, a DARPA program launched in 2025 to build an AI co-author that proposes and proves useful mathematical abstractions.

How does HAYAT relate to autonomous research?
HAYAT is Mohsin Khan Shaik's open-source local AI assistant. It runs cited background research, learns topics it couldn't answer, tracks how accurate its predictions are, and only reports actions that tool results prove.

Glossary

Self-driving lab: A laboratory where AI plans experiments and robots run them in a closed loop.

Provenance: A record of where a claim, dataset or result came from.

Neuro-symbolic AI: Systems that combine neural networks with formal logic or symbolic reasoning.

Drug repurposing: Finding new uses for drugs that are already approved for other conditions.

References

1. Singularity Hub — What the rise of AI scientists may mean for human research (Feb 2026)
   https://singularityhub.com/2026/02/20/what-the-rise-of-ai-scientists-may-mean-for-human-research/
2. DARPA — expMath: Exponentiating Mathematics
   https://www.darpa.mil/research/programs/expmath-exponential-mathematics
3. DARPA — Accelerating mathematical discovery with AI (2025)
   https://www.darpa.mil/news/2025/math-ai-tomorrows-breakthroughs
4. University of Maryland — UMD to lead DARPA-funded effort to accelerate mathematical discovery with AI
   https://today.umd.edu/umd-to-lead-darpa-funded-effort-to-accelerate-mathematical-discovery-with-ai

5. DARPA — MATHBAC program
   https://www.darpa.mil/research/programs/mathbac
6. Google DeepMind — AlphaEvolve: a Gemini-powered coding agent for designing advanced algorithms (May 2025)
   https://deepmind.google/discover/blog/alphaevolve-a-gemini-powered-coding-agent-for-designing-advanced-algorithms/
7. DARPA — DICE: Decentralized AI through Controlled Emergence
   https://www.darpa.mil/research/programs/decentralized-artificial-intelligence-through-controlled-emergence
8. HAYAT on GitHub (README, architecture and test counts)
   https://github.com/mohsinkhanshaik/HAYAT
9. Scientific American — New Google AI tackles complex math and science
   https://www.scientificamerican.com/article/new-google-ai-chatbot-tackles-complex-math-and-science/

Online: https://mohsinkhanshaik.com/research/autonomous-discovery · Facts are cited; outlooks and "My take" are the author's opinion.