Proceedings of the Faculty of Technical Sciences, University of Novi Sad

ФТН ЗБОРНИК

State-of-the-Art Review
UDK: 4.8
DOI: https://doi.org/10.24867/36WS022611

AI Scientists' View of the AI Scientists

Dubravko Ćulibrk[1,2]

[1] *Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia,* [2] *Jožef Stefan Institute, Ljubljana, Slovenia*

[1]dculibrk@uns.ac.rs

Abstract – *For a long time, "AI scientist" meant a person who builds artificial intelligence systems. Lately, the term has also come to mean software systems that try to carry out the whole research cycle by themselves, from reading the literature to writing and reviewing the paper. In 2026 such systems moved from curiosity to headline: a fully AI-written paper passed peer review and was described in Nature, and in September a swarm of some ten thousand AI agents produced a computer-verified proof that the three-dimensional Navier–Stokes equations can develop singularities, settling one of the Millennium Prize Problems and provoking a sharp reaction from the mathematical community. This article provides an informal overview of the field from the point of view of a researcher who has worked in machine learning, computer vision and multimedia since 2002. It explains how current AI scientist systems are organised, how much attention and public funding the topic now receives, including the European RAISE initiative, looks at the Navier–Stokes episode as a case study, and summarises what the evidence says about the strengths and weaknesses of these systems. The systems are fast, tireless and very effective where results can be checked automatically, but they remain weak at choosing questions and at producing understanding rather than answers. The article closes with the question many researchers are now asking each other: how do we take it from here?*

Keywords: artificial intelligence, AI scientist, automated scientific discovery, AI agents, Navier–Stokes, research practice

1. INTRODUCTION

The domain of machine learning research has undergone an amazing tranformation over the last 25 years. At the turn of the century, the scientist did everything: chose the problem, designed the features, wrote the code, ran the experiments and wrote the paper. Interestingly enough, many reseach topics were as relevant then as they are now. E.g., trying to teach a machine to optimize the code and increase the speed of its execution for another machine [1]. Although a quarter of a century has passed, this remains one of the most active topics of research in the domain of computer engineering to this day [2]. There are, however, some sharp differences to how we approach the problem now and then. The learning algorithm used to be a small, carefully supervised component in the middle. Over the following two decades, that component grew. More than a decade ago, deep learning took over feature design, then the foundation models took over much of the modelling, and coding assistants took over a good part of the programming. Each step moved the boundary between what the researcher does and what the machine does, and each time the boundary seemed to settle in a new, comfortable place.

In 2026 it stopped settling. "AI scientist", which used to mean a person doing research in the domain of AI, now also names software systems that attempt to do research themselves. The best known is literally called *The AI Scientist*, and in March 2026 its description appeared in *Nature* [3], a journal its human namesakes very much hope to publish in. In September, OpenAI announced that a swarm of about ten thousand AI agents had settled the Navier–Stokes Millennium Prize Problem [4], and more than two dozen Fields medallists responded with a public warning about where this is taking mathematics [5]. Nor is this only something human AI scientists read and about. As these lines are being written, an AI agent is running remote sensing experiments for the author on another computer, trying to improve detection of organic carbon in the soil from satellite imagery. The title of this article is therefore meant in both senses: one kind of AI scientist looking at the other, while the other is busy working for him.

Last month two students from our research group attended MICCAI 2026, the premier conference on medical image computing [6]. The main topic of conversation among the scientists there was not a particular method or dataset, but a question: how do we take it from here? This article is an attempt at an answer. Section 2 provides an overview of the journey that got us here, Section 3 examines how much attention, and investment, the

topic now attracts, Section 4 describes what the systems actually do, Section 5 takes a look at the Navier–Stokes episode, Section 6 reviews what measured results and benchmarks show about the performance of these systems, while Section 7 focuses on the future. Fittingly, this article was itself written with the help of an AI agent; the declarations at the end describe how.

2. FROM INSTRUMENTS TO COLLEAGUES

The idea that discovery could be automated is about as old as AI. In the late 1970s, Langley, Simon and colleagues built programs such as BACON that rediscovered physical laws, including Kepler's third law, from tables of measurements [7]. Kepler needed years to find that law; BACON needed seconds, provided that someone had already chosen the right variables. In 2009, King and colleagues described Adam, a "robot scientist" that formed hypotheses about yeast genetics, ran the experiments with laboratory automation and interpreted the results [8]. Kitano later proposed the "Nobel Turing Challenge": an AI system capable of Nobel-level discoveries by 2050 [9].

For most of the last decade, however, AI entered science as an instrument. AlphaFold [10], recognised by the 2024 Nobel Prize in Chemistry [11], solved a problem that scientists had defined, but it did not choose the problem or decide what to do with its predictions. Large language models (LLMs) changed this because they handle the "soft" parts of research that earlier systems could not: reading papers, writing code, explaining results and arguing for an idea. Put an LLM in a loop in which it can call tools, run programs and inspect their output, and it becomes an *agent*. From an agent to an AI scientist is just a short step.

3. WHY EVERYONE IS PAYING ATTENTION

AI for science is no longer a niche topic, and several indicators show it. In 2024, two Nobel Prizes went to AI: the Physics prize to Hopfield and Hinton for foundational work on artificial neural networks [12], and half of the Chemistry prize to Hassabis and Jumper for AlphaFold [11]. The 2026 Stanford AI Index added, for the first time, a separate chapter on science. It counts about 80,150 AI-related publications in the natural sciences in 2025, 26% more than a year earlier, and notes that AI now appears in roughly one in fifteen papers in these fields, compared with less than 1% in 2010 [13]. The submission figures of the large AI conferences, discussed in Section 7, tell the same story.

Governments have noticed as well. In November 2025, the United States launched the "Genesis Mission", a national initiative led by the Department of Energy that aims to build a shared platform of supercomputers, AI models and scientific data for AI-driven discovery, and that its proponents openly compare in ambition to the Manhattan Project [14]. A month earlier, the European Commission adopted "A European Strategy for Artificial Intelligence in Science" [15]. Its centrepiece is RAISE, the Resource for AI Science in Europe, conceived as a virtual European institute that pools computing, data, talent and research funding [15], [16]. The RAISE pilot was launched in November 2025 and receives EUR 108 million under the 2026–2027 Horizon Europe work programme, on top of initial funding in 2025 [15], [16]. It covers networks of excellence, doctoral networks on AI in science, scientific foundation models and, notably, laboratory automation [15]. Horizon Europe will also invest up to EUR 600 million to secure researchers dedicated access to computing capacity in the planned AI Gigafactories, and the Commission will seek to double Horizon Europe's yearly investment in AI, including in AI for science, by 2028 [15]. For context, Horizon Europe has supported AI with more than EUR 8 billion since 2021 [15].

The European plan has received a mixed reception. Critics point out that the amounts are small compared with what private companies spend, that most of the money will only become available towards the end of the decade, and that multi-year project funding is a poor fit for a field that changes direction every few months [17]. The Navier–Stokes episode, in which a single company reportedly spent about a tenth of the entire RAISE pilot budget in one week, illustrates the point. Still, for researchers in Serbia, which is associated to Horizon Europe, these calls are one of the few realistic ways to obtain the computing and the collaborations that this kind of research requires.

Computing is where the European effort is of most tangible use to human (AI) scientists. Through the EuroHPC Joint Undertaking, the EU has so far elected to build 19 AI Factories, supercomputing centres optimised for AI that give researchers, start-ups and industry access to computing, data and expertise, together with 13 smaller AI Factory Antennas that connect countries that have no AI factory to these institutions [18]. The next step is a network of up to seven AI Gigafactories, very large facilities built with substantial private co-investment, for which the call for tenders closes in November 2026 [19]. Of the countries formerly part of Yugoslavi, only Slovenia is building an AI Factory at the momen, which will install in Maribor a supercomputer with about 10 exaflops of mixed-precision performance for AI [20]. Serbia is extending its own AI infrastructure, but has a foothold in the AI Factory community as well: the Serbian AI Factory Antenna (SAIFA), run by a consortium of eight Serbian institutions, links national infrastructure with the Pharos AI Factory in Greece and the IT4LIA AI Factory in Italy [18], [21].

For scientists who want to bring synthetic AI scientist colleagues into their research, these facilities matter in three ways. First, they provide computing time through open access calls, at a scale no single university in the region could afford, and computing is exactly what agent-driven experiments consume. Second, they come with people: support teams that help researchers move from a prototype on a workstation to models trained or fine-tuned at scale, and that help prepare data in areas such as health, language and the environment. Third, they keep data and models within European infrastructure and rules, which matters for sensitive data, such as medical images, that cannot simply be sent to a commercial cloud. For researchers in Serbia, SAIFA and the links to neighbouring AI Factories offer a realistic way in. The challenge, as with RAISE, will be to make access fast and simple enough for a field that moves in weeks rather than years.

4. WHAT AN AI SCIENTIST DOES TODAY

Fig. 1 shows the research cycle as most current systems implement it. A human gives a broad direction, and the system then reads the literature and proposes hypotheses, ranks them (usually by asking the model itself to judge novelty and feasibility), runs experiments, which today mostly means writing and executing code, analyses the results, writes a report or a complete paper, and finally reviews its own work with a separate reviewer agent before starting another round.

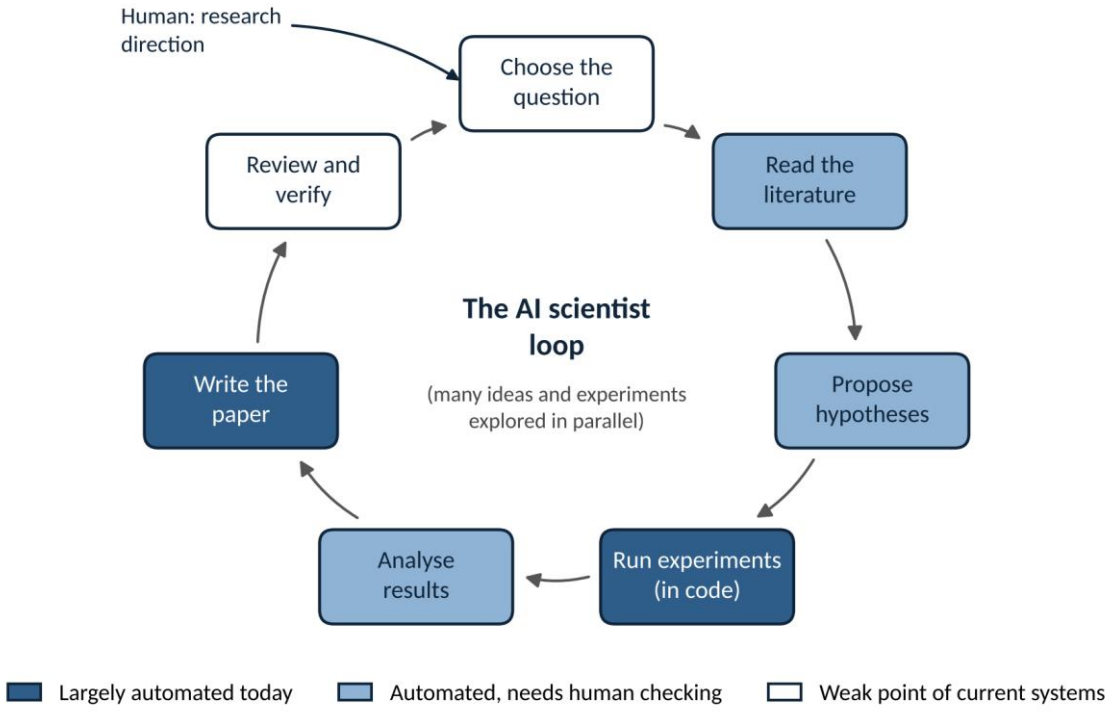


Figure 1. The research cycle as implemented by current AI scientist systems. Shading indicates, informally, how well each stage is handled today.

Table 1 lists representative systems. Google's AI co-scientist [22] concentrates on generating and debating hypotheses and leaves experiments to human laboratories. Coscientist [23] connects the language model to laboratory hardware. AlphaEvolve [24] lets the model propose programs that an automatic evaluator scores, so a solution survives only if it is measurably better. The AI Scientist [3] and Kosmos [25] aim at the whole cycle. Two ingredients are common to almost all of them. The first is *search*: many ideas and experiments are tried in parallel and the most promising are kept, as in the tree of experiments used by The AI Scientist [3]. The second is *memory*, such as the shared "world model" in Kosmos [25], which lets the system keep the thread over hours of work and hundreds of agent runs.

Table 1. Representative AI systems for scientific research.

System (year)	Domain	What it automates	Reported result
Adam (2009) [8]	Yeast genetics	Hypotheses and robotic laboratory experiments	Identified genes for several “orphan” enzymes
Coscientist (2023) [23]	Chemistry	Planning and running experiments on lab hardware	Optimised catalytic coupling reactions
AlphaEvolve (2025) [24]	Mathematics, algorithms	Program search with automatic evaluation	Faster 4×4 complex matrix multiplication
AI co-scientist (2025) [22]	Biomedicine	Generating and debating hypotheses	Drug-repurposing candidates supported in vitro
Kosmos (2025) [25]	Data-driven science	Literature search and data analysis in long runs	79.4% of report statements judged accurate
The AI Scientist (2024–2026) [3]	Machine learning	Idea, experiments, paper and review	One paper accepted at an ICLR 2025 workshop
OpenAI agent swarm (2026) [4]	Mathematics	Proof search with formal verification in Lean	Blow-up for 3D Navier–Stokes, pending peer review

5. THE NAVIER–STOKES CONUNDRUM

The Navier–Stokes equations describe how fluids move, from blood in arteries to air over a wing, and most engineering students meet them. Whether their solutions in three dimensions always stay smooth, or can break down in finite time, is one of the seven Millennium Prize Problems set by the Clay Mathematics Institute. On 1 September 2026, OpenAI launched a large multi-agent effort on open mathematical problems. On 8 September it announced that roughly 10,000 agents, working for 88 hours and exchanging about 2.7 million messages, had produced a proof that solutions *can* blow up, together with a formal version of the proof checked in the Lean proof assistant [4], [26]. OpenAI described the cost only as multimillion-dollar; figures of around 15 million US dollars were reported [26], [27].

The reaction was anything but a simple celebration. About twelve hours before the announcement, Tristan Buckmaster and Levent Alpöge had announced closely related results, and Buckmaster questioned whether OpenAI had benefited from his own use of its coding tools; the company stated that it could not rule out that de-identified usage data had helped improve its models [28]. Mathematicians also pointed out that the agents’ strategy built directly on years of human work by Córdoba, Martínez-Zoroa and others [4]. Buckmaster described the first LLM-generated proof he received as the most horrendous he had ever read [4]. The Clay Institute said the problem had “apparently been settled” but that any prize depends on peer-reviewed publication [27]. Most tellingly, a group of Fields medallists argued that the goals of AI companies and of mathematicians are severely misaligned, because a famous problem is valued as a proxy for conceptual understanding, not as an answer in itself [5].

The episode is a near-perfect summary of the field. The question was chosen by humans a quarter of a century ago. The route ran through human ideas. The machines contributed enormous, brute-force search, and the result can be trusted because a proof checker, not a reviewer, verified it. What is missing, for now, is the part mathematicians care about most: understanding why the solution blows up, in a form that one person can explain to another. And the price tag shows who can play this game at this scale.

6. WHAT THE EVIDENCE SAYS

Away from the headlines, the measured results tell a consistent story. The AI Scientist paper accepted in 2025 was one of three submissions, at a workshop with an acceptance rate of about 70%, and its authors state that none would meet main-conference standards [3]. An independent evaluation of the first version found that 42% of its experiments failed because of coding errors and that some papers contained hallucinated numbers [29]. LLM-generated research ideas were rated more novel than those of human experts, but slightly less feasible [30]. One in five Kosmos statements is not accurate [25], and in science the wrong fifth may be the one that

matters. At Agents4Science, a conference where AI had to be the first author, reviewers found accepted papers technically sound but often lacking an interesting question [31], [32]. Systematic benchmarks agree. According to the AI Index, the best agent answered 38.8% of questions on the PaperArena benchmark, against 83.5% for PhD-level experts, and frontier models reproduced fewer than 20% of the astrophysics results in ReplicationBench [13]. Three observations follow.

First, these systems are strongest where results can be checked automatically. A tireless generator of guesses combined with a reliable checker, whether a benchmark, a simulator or a proof assistant, is extremely powerful, as AlphaEvolve and the Navier–Stokes proof show. When the checker is another language model, the loop is much weaker, because the reviewer shares the author's blind spots.

Second, the hardest part of research is deciding what is worth doing. Current systems optimise within a direction given by a human. They are good at well-defined improvements and poor at recognising that a question is boring, that an assumption is wrong, or that an unexpected result is more interesting than the planned one.

Third, producing a convincing output is easier than producing understanding. LLMs write fluently, so weak results can look strong, and even correct results can arrive without insight. Messeri and Crockett warn that AI tools can create "illusions of understanding" in science [33]; for systems that write their own papers, the risk is built in.

7. HOW DO WE TAKE IT FROM HERE?

Back to the question from MICCAI. The honest answer is neither "AI will replace us" nor "nothing changes". A few things seem likely.

Progress will be fastest where verification is cheap: mathematics with proof assistants, algorithm design, simulation-based science and data analysis with known ground truth. Medical imaging sits in between. Segmentation accuracy can be measured automatically, clinical value cannot.

Agents will move from computers into physical laboratories, continuing the line from Adam [8] to Coscientist [23]. This is no longer limited to large industrial labs. At a lecture at the Jožef Stefan Institute in Ljubljana this week, Matej Humar, one of the rising stars of photonics, described how two postdocs in his group are working on using AI agents to automate their research, including physical experiments with lasers. Here the bottleneck is less the intelligence of the agent than the cost and flexibility of the hardware.

The productive unit of research will be a human scientist with a team of fast, capable and occasionally careless AI assistants. At the AutoML conference held just last week, a colleague from Stanford University remarked that a postdoc used to be given two Master students to help with the work, and is now given a budget for AI agents instead. The human still chooses the problem, decides which results to trust and takes responsibility for what is published, which is why most journals do not accept AI systems as authors [34].

Credit, priority and data will become research-ethics issues in their own right. The Navier–Stokes dispute shows that when the same companies provide researchers with their tools and compete with the for results, questions about who saw what, and when, are no longer hypothetical.

Compute will shape who can do frontier science. An experiment that costs millions of dollars in a week is out of reach of any university in the region. The advantage of researchers in the region lies in good questions, domain knowledge, data that others do not have, and the ability to check results critically.

There is a more immediate consequence, and it is visible in the numbers. Table 2 compares submissions to several major AI conferences in 2026 with the previous year. The International Conference on Machine Learning (ICML) doubled in a single year [35]. For its 2027 edition, with the deadline in September 2026, the International Conference on Learning Representations (ICLR) registered more than 60,000 abstracts, about three times its previous total [36]. Not all of this growth comes from AI-written papers; the field itself is growing quickly. Still, the trend accelerated just as writing a plausible paper became cheap. At Neural Information Processing Systems (NeurIPS) 2026, the leading conference in the domain, the chairs desk-rejected 18% of submissions to the position-paper track as excessively AI-generated [37].

Table 2. Submissions to selected AI conferences, previous and current cycle.

Conference	Previous cycle	Current cycle	Change
ICLR (2026 → 2027) [36]	≈19,500	>60,000 (abstracts)	≈ ×3
ICML (2025 → 2026) [35]	12,107	23,918	+98%
CVPR (2025 → 2026) [38]	12,974	16,092	+24%

For peer review, this means more weight on released code and data, reproducibility and the significance of

the question, and less on polish. For engineering education, and for faculties such as the Faculty of Technical Sciences in Novi Sad, the message has two parts. Students must learn to use these tools well, because their employers and research groups will expect it. At the same time, the skills that gain value are exactly those the tools lack: formulating a good question, designing an experiment that can actually fail, checking a result independently, and explaining *why* something is true. The Leskovec remark also hides a problem. Master students used to learn research by doing the routine parts of it next to a postdoc. If agents take over those parts, universities have to find another way to train the next generation, or there will be no one left to supervise the agents. A course that only teaches students to produce a report is teaching a skill that has just been automated.

8. CONCLUSION

The AI scientist in its new, software sense is no longer a thought experiment. It writes papers that pass review, and with enough compute and a good formal checker it can settle problems that resisted people for generations. It is also much better at carrying out research than at choosing it, and better at producing answers than understanding. After more than twenty years of watching the boundary move between what the researcher does and what the machine does, one finds that this time the boundary has moved to the part of the job that was always the most important: asking the right question and knowing why the answer is right. That is where the research community should take it from here.

FUNDING

This research received no external funding.

DECLARATIONS

Conflict of interest: The author declares no conflict of interest.
Data availability: Not applicable; no new data were created or analysed in this article.
Author contributions: D.Ć. conceived and wrote the article.
Ethical approval: Not applicable.
Declaration of AI use: In keeping with its subject, this article was prepared with substantial help from an AI agent, Claude (Anthropic). The author defined the topic, title, perspective and argument, and supplied all personal observations and anecdotes. The agent searched for and summarised recent literature and news reports, drafted the text based on the author's instructions, compiled the tables and references, produced Figure 1, and formatted the manuscript according to the Proceedings template. The author reviewed and edited the text, checked the cited sources, and takes full responsibility for the content. No AI system is listed as an author.

REFERENCES

The author is responsible for the originality, quality and accuracy of the results presented in the paper.

[1] D. Ćulibrk, "Jedno rešenje izbora suboptimalne strategije sažimanja mašinskog koda jedne klase procesora primenom genetskog algoritma [A solution for selecting a suboptimal machine-code compaction strategy for a class of processors using a genetic algorithm]," M.Sc. thesis, Fac. Tech. Sci., Univ. Novi Sad, Novi Sad, Serbia, 2003 (in Serbian).

[2] Y. Yu *et al.*, "Towards automated kernel generation in the era of LLMs," in *Proc. 35th Int. Joint Conf. Artif. Intell. (IJCAI)*, Bremen, Germany, 2026, pp. 8069–8077, doi: 10.24963/ijcai.2026/895.

[3] C. Lu, C. Lu, R. T. Lange, Y. Yamada, S. Hu, J. Foerster, and J. Clune, "Towards end-to-end automation of AI research," *Nature*, vol. 651, pp. 914–919, Mar. 2026, doi: 10.1038/s41586-026-10265-5.

[4] K. Kakaes, "AI has solved one of math's $1 million Millennium Prize Problems," *Quanta Magazine*, Sep. 8, 2026. [Online]. Available: https://www.quantamagazine.org/ai-has-solved-one-of-maths-1-million-millennium-prize-problems-20260908/

[5] "A severe misalignment of AI in mathematics," declaration, Sep. 11, 2026, doi: 10.5281/zenodo.22737750. [Online]. Available: https://mathandai.org

[6] N. Jovišić, M. Škipina, N. Dall'Asen, and D. Ćulibrk, "MIL-PF: Multiple instance learning on precomputed features for mammography classification," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv. (MICCAI)*, Strasbourg, France, 2026, pp. 452–462.

[7] P. Langley, H. A. Simon, G. L. Bradshaw, and J. M. Zytkow, *Scientific Discovery: Computational Explorations of the Creative Processes*. Cambridge, MA, USA: MIT Press, 1987.

[8] R. D. King *et al.*, "The automation of science," *Science*, vol. 324, no. 5923, pp. 85–89, Apr. 2009.

[9] H. Kitano, "Nobel Turing Challenge: Creating the engine for scientific discovery," *npj Syst. Biol. Appl.*, vol. 7, Art. no. 29, 2021.

[10] J. Jumper *et al.*, “Highly accurate protein structure prediction with AlphaFold,” *Nature*, vol. 596, pp. 583–589, Aug. 2021.
[11] The Royal Swedish Academy of Sciences. “The Nobel Prize in Chemistry 2024.” Accessed: Oct. 7, 2026. [Online]. Available: https://www.nobelprize.org/prizes/chemistry/2024/summary/
[12] The Royal Swedish Academy of Sciences. “The Nobel Prize in Physics 2024.” Accessed: Oct. 7, 2026. [Online]. Available: https://www.nobelprize.org/prizes/physics/2024/summary/
[13] Stanford Institute for Human-Centered Artificial Intelligence, “Science,” in *The AI Index 2026 Annual Report*, ch. 5. Stanford, CA, USA: Stanford University, 2026. [Online]. Available: https://hai.stanford.edu/ai-index/2026-ai-index-report/science
[14] The White House, “Launching the Genesis Mission,” Executive Order 14363, Washington, DC, USA, Nov. 24, 2025.
[15] European Commission, “A European strategy for artificial intelligence in science: Paving the way for the Resource for AI Science in Europe (RAISE),” Communication from the Commission to the European Parliament and the Council, COM(2025) 724 final, Brussels, Belgium, Oct. 8, 2025. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52025DC0724
[16] European Commission. “RAISE: Resource for AI Science in Europe.” Accessed: Oct. 7, 2026. [Online]. Available: https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/european-ai-science-strategy/raise-resource-ai-science-europe_en
[17] D. Matthews, “The EU’s AI in science plan gets a mixed reception,” *Science|Business*, Nov. 6, 2025. [Online]. Available: https://sciencebusiness.net/news/ai/eus-ai-science-plan-gets-mixed-reception
[18] EuroHPC Joint Undertaking, “EuroHPC JU selects AI Factory Antennas to broaden AI Factories initiative,” press release, Luxembourg, Oct. 13, 2025. [Online]. Available: https://www.eurohpc-ju.europa.eu/eurohpc-ju-selects-ai-factory-antennas-broaden-ai-factories-initiative-2025-10-13_en
[19] EuroHPC Joint Undertaking, “The EuroHPC Joint Undertaking launches the AI Gigafactories call,” press release, Luxembourg, Jul. 30, 2026. [Online]. Available: https://www.eurohpc-ju.europa.eu/eurohpc-joint-undertaking-launches-ai-gigafactories-call-2026-07-30_en
[20] SLING. “Slovenian AI Factory (SLAIF).” Accessed: Oct. 7, 2026. [Online]. Available: https://www.sling.si/en/slaif/
[21] HPC Serbia. “Serbian Artificial Intelligence Factory Antenna.” Accessed: Oct. 7, 2026. [Online]. Available: https://hpc.rs/news/serbian-artificial-intelligence-antenna-factory
[22] J. Gottweis *et al.*, “Accelerating scientific discovery with Co-Scientist,” *Nature*, vol. 655, pp. 487–496, 2026, doi: 10.1038/s41586-026-10644-y.
[23] D. A. Boiko, R. MacKnight, B. Kline, and G. Gomes, “Autonomous chemical research with large language models,” *Nature*, vol. 624, pp. 570–578, Dec. 2023.
[24] A. Novikov *et al.*, “AlphaEvolve: A coding agent for scientific and algorithmic discovery,” 2025, *arXiv:2506.13131*.
[25] L. Mitchener *et al.*, “Kosmos: An AI scientist for autonomous discovery,” 2025, *arXiv:2511.02824*.
[26] C. Franzen, “OpenAI solves longstanding math problem with 10,000-agent swarm — but can’t rule out benefitting from a researcher’s private Codex data,” *VentureBeat*, Sep. 8, 2026.
[27] Wikipedia. “Navier–Stokes priority controversy.” Accessed: Oct. 7, 2026. [Online]. Available: https://en.wikipedia.org/wiki/Navier%E2%80%93Stokes_priority_controversy
[28] M. Mills, “OpenAI’s historic math solution overshadowed by credit controversy,” *Axios*, Sep. 8, 2026.
[29] J. Beel, M.-Y. Kan, and M. Baumgart, “Evaluating Sakana’s AI Scientist: Bold claims, mixed results, and a promising future?” 2025, *arXiv:2502.14297*.
[30] C. Si, D. Yang, and T. Hashimoto, “Can LLMs generate novel research ideas? A large-scale human study with 100+ NLP researchers,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Singapore, 2025.
[31] F. Bianchi, O. Queen, N. Thakkar, E. Sun, and J. Zou, “Exploring the use of AI authors and reviewers at Agents4Science,” 2025, *arXiv:2511.15534*.
[32] Science News. “A conference just tested AI agents’ ability to do science.” Accessed: Oct. 7, 2026. [Online]. Available: https://www.sciencenews.org/article/science-conference-test-ai-agents
[33] L. Messeri and M. J. Crockett, “Artificial intelligence and illusions of understanding in scientific research,” *Nature*, vol. 627, pp. 49–58, Mar. 2024.
[34] “Tools such as ChatGPT threaten transparent science; here are our ground rules for their use,” *Nature*, vol. 613, p. 612, Jan. 2023.
[35] CS Conference Statistics. “ICML acceptance rate and submission statistics.” Accessed: Oct. 7, 2026. [Online]. Available: https://csconfstats.xoveexu.com/conferences/icml/
[36] Reviewer3. “ICLR went from 490 submissions to 62,000 in ten years.” Accessed: Oct. 7, 2026. [Online]. Available: https://reviewer3.com/evidence/iclr-submissions-2027

[37] NeurIPS Program Chairs. “AI-generated papers in the NeurIPS 2026 position paper track.” Accessed: Oct. 7, 2026. [Online]. Available: https://blog.neurips.cc/2026/06/02/ai-generated-papers-in-the-neurips-2026-position-paper-track/
[38] CVPR 2026, “CVPR 2026 fields 16,000+ paper submissions on technical advances in AI,” press release, May 28, 2026. [Online]. Available: https://www.prnewswire.com/news-releases/cvpr-2026-fields-16-000-paper-submissions-on-technical-advances-in-ai-302784529.html

AUTHOR BIOGRAPHY



Dubravko Ćulibrk is a Full Professor at the Faculty of Technical Sciences, University of Novi Sad, Serbia, and the founding director of the Institute for Artificial Intelligence Research and Development of Serbia. He is also a Visiting Researcher at the Jožef Stefan Institute in Ljubljana, Slovenia. He has worked in machine learning, computer vision and multimedia for nearly 25 years, and his current research covers Computer Vision, multimodal AI and Remote Sensing. He is a member of Serbia’s National AI Council and an expert of the Global Partnership on Artificial Intelligence, and serves as Chief Science Officer of SmartCloudFarming GmbH, a company developing AI-based soil intelligence. Contact: dculibrk@uns.ac.rs.