

Trustworthy AI-Enabled Mathematics and Scientific Computing

Scientific directions and opportunities for collaboration

Workshop link (including all videos): [PDEs, Verification, Mathematics, and Algorithms in the Age of AI · September 17–18, 2026](#)

Jeremy Avigad (ICARM)

Javier Gomez-Serrano (Brown University)

Jeffrey Hittinger (Lawrence Livermore National Laboratory)

Motoko Kotani (RIKEN)

Dima Shlyakhtenko (UCLA and IPAM)

Terence Tao (UCLA and IPAM)

Tatiana Toro (Simons Laufer Mathematical Sciences Institute)

Michio Yoshikawa (Tohoku University)

An emerging scientific landscape

AI is expanding the capacity to generate mathematical arguments, numerical algorithms, and scientific software. The IPAM workshop explored how these capabilities might translate into reliable scientific knowledge. A central opportunity is to connect **AI for mathematics and science** with **mathematics for reliable, efficient AI**. Potential advances include certified solutions of partial differential equations (PDEs), reconstruction of governing laws from observations, and wider access to sophisticated numerical methods.

The vision of the U.S. Government's [Genesis Mission](#) is an integrated ecosystem of resources that empower scientists and engineers with autonomous discovery, design, experimentation, and manufacturing. To enable such autonomy requires confidence in the capabilities of the ecosystem, which are continuously evolving. Consider, for instance, an engineer developing a predictive model for design optimization from a collection of experimental observations. The workflow may involve model discovery (identifying a set of closure models for PDE governing equations), the generation of computer code that adds a discrete approximation to an existing simulation code, running an ensemble of simulations to validate the model relative to the

experimental data, and so on. Throughout this workflow, questions of correctness exist that are fundamentally issues for mathematics, including the well-posedness of the discovered model, the stability and consistency of the discrete model, the accuracy of computed approximate solutions relative to true solutions of the governing equations, and the ability of the model to accurately represent reality.

Three questions provide a useful organizing framework: **Is the mathematical claim valid? Does the computation approximate the stated problem with controlled error? Does the problem adequately represent the physical system?** Formal proofs, numerical analysis, and empirical validation address different parts of this framework, and all three can be potentially accelerated by judicious use of Large Language Models and other machine learning techniques. Proof assistants, such as Lean and Rocq, check the formal specification they receive; determining whether that specification captures the scientific intention remains a distinct intellectual task. This distinction underlies the scientific and organizational opportunities outlined below.

[Mathematical sciences research institutes](#) offer a natural setting for closer collaboration between [NSF](#), which supports the institute network, and DOE through the [Genesis Mission](#), which is building an AI-enabled platform linking computing, experiments, and scientific data. Institute programs could help close critical communication gaps among mathematicians, AI researchers, verification specialists, and laboratory scientists by creating shared language around correctness and translating application challenges into mathematical questions.

The workshop demonstrated that there is an opportunity to build on these models through thematic programs, laboratory–university residencies, joint training, technical assistance, and continuing working groups. NSF’s support for open mathematical communities and DOE’s mission problems, facilities, data, and computing are complementary strengths. Closer connections centered on institutes could help develop the scientific communities needed for the changes discussed here. Shared specifications, benchmarks, reusable libraries, and durable laboratory collaborations would be valuable outcomes, providing human and intellectual infrastructure alongside Genesis’s technological platform. The workshop also suggested a concrete basis for US–Japan collaboration; the US–Japan AI-for-science partnership announced in June 2026 provides a broader context for such exchange based on longstanding international collaborations between two countries.

1. Shared foundations for trustworthy computation

Reusable mathematical and software infrastructure emerged as a central enabling need during the workshop discussions. Participants identified gaps in formal PDE analysis and in the layers connecting equations (and model discovery) to executable solvers, including discretization, mesh geometry, numerical schemes, and error estimates. Individual verified examples offer

valuable starting points; their wider usefulness depends on developing common abstractions and interfaces.

[Andrew Appel](#)'s component-by-component finite-element project illustrates one approach: connecting mathematical approximation results to floating-point accuracy and proofs about executable code. Composition remains a substantial challenge, since the guarantees of one component must meet the assumptions of another. Lean and Rocq offer opportunities to connect mathematical and software-verification expertise across communities. Both are rapidly developing, with strong support for analysis in Lean and strong support for numerics in Rocq.

Progress along these lines depends on collaboration among PDE analysts, numerical analysts, formal-methods researchers, and research software engineers. Proof assistant library curation, documentation, version compatibility, and long-term maintenance are integral to scientific usability, particularly for simulation codes that remain in service for decades.

2. From numerical candidates to checkable guarantees

A promising approach separates discovery from checking. A numerical solver or neural network proposes an approximate solution; analysis and rigorous computation then establish a precise guarantee. [Kazuaki Tanaka](#) presented learned sub- and super-solutions that enclose true solutions under suitable assumptions. [Bhavik Mehta](#) described formalizing the analysis and arithmetic needed to turn approximate computations into theorems. These approaches make it possible to use powerful search methods that lead to trusted results without requiring the process itself be fully verified.

An important question is how certification can become practical for scientifically significant PDEs, parameter ranges, and dimensions. Relevant directions include residual and stability estimates, bounds on omitted terms, efficient interval and floating-point arithmetic, and dependable certificate checkers. Verifying a reusable solver and checking a particular computed result provide complementary routes, with different costs and trust assumptions.

Carefully chosen demonstration problems could clarify these tradeoffs. Comparisons with conventional methods, such as a posteriori error estimation, order verification, benchmarking, regression testing, etc., would be most informative when they include applicability, error guarantees, remaining assumptions, and total discovery-and-certification cost. Such work brings together mathematical and software expertise, CPU resources for checking, GPU resources where learning is involved, and close engagement with application scientists.

3. Scientific understanding and validation of AI-enabled methods

Realizing these possibilities involves more than producing answers quickly. [Allison Baker](#)'s climate-model examples showed how small changes in floating-point evaluation can interact with code discontinuities to alter simulation statistics. [Matthew Sottile](#) described the many stages between a physical model and a computational prediction: discretization, meshes, solvers, arithmetic, and parallel software. Understanding how assumptions and errors propagate across these stages becomes increasingly important as AI accelerates scientific development.

Mathematics can contribute to the design of AI methods as well as to their assessment. [Takaharu Yaguchi](#)'s structure-preserving learning and [Robert Joseph George](#)'s verified scientific machine-learning work illustrate complementary approaches: incorporating mathematical structure into architectures and making computational properties checkable. Stability, generalization across physical regimes, mixed precision, and the transfer of guarantees to actual hardware remain important research questions. Structure preservation alone does not establish complete accuracy or physical validity.

[Koji Hashimoto](#) emphasized the inverse scientific problem of recovering interpretable governing principles from observations. His work raises questions about which laws the data can distinguish, how noise and uncertainty affect inference, and how physical constraints inform learning. The Navier–Stokes panel discussion similarly distinguished exceptional mathematical constructions from physically typical behavior. Understanding mechanisms, stability, and realistic regimes remains valuable independently of the resolution of individual conjectures.

For complex applications, formal verification complements statistical testing and experimental validation. [Baker](#)'s ensemble consistency tests detect changes without demanding bitwise-identical simulations, but do not determine which version is physically correct. Connecting such diagnostics to error analysis and model validation offers a productive direction. Curated observations, challenging regression tests, heterogeneous hardware, and explicit numerical specifications would help make comparisons meaningful. AI agents directing scientific workflows introduce a related question: how to assess the reliability of the overall process, beyond its individual computations.

A workshop panel also discussed verification and validation in PDEs and machine learning models. Recent progress in PDE formalization, the development of reusable libraries for PDE analysis (Wuyang Chen and Jacob Bedrossian), and Lean-based formalization technologies (Sho Sonoda) were introduced. Taiji Suzuki also highlighted the potential of mathematical formalization for the study of intelligent systems. However, a significant gap remains between informal and formal proofs. Bridging this gap may be crucial not only for advancing machine verification, but also for making machine-generated proofs accessible to human understanding. Developing methods to connect informal mathematical reasoning with formal proofs therefore

represents an important new research challenge at the intersection of mathematics, formalization, and machine learning.

4. Understanding, access, and the next generation of researchers

The workshop repeatedly distinguished machine-checked truth from human understanding. Large generated proofs may be difficult to maintain or use as sources of new ideas. [Terence Tao](#)'s network of formal results illustrates the importance of curated statements, explicit dependencies, and manageable interfaces to complex proofs. Explanation, refactoring, independent review, and connections between ordinary mathematical language and formal statements are therefore substantive research contributions.

Graduate and postdoctoral training, summer schools, institute programs, and sustainable software careers can help develop the necessary expertise. Shared access to AI tools can broaden participation, especially at smaller institutions, and perhaps this is an important role for the Genesis Mission: to provide AI “facilities” access to great ideas much the way DOE and NSF maintain other user facilities. An important educational question is how to combine AI and proof assistant tools with the development of independent reasoning and deep subject knowledge. Recognition of collaborative, expository, and infrastructure-building work would help align scientific practice with these changing needs.

5. Mathematical sciences institutes as bridges among communities

The [NSF Mathematics Institutes](#) could serve an important role in bridging between diverse technical communities as AI affects all of them. IPAM's experience provides concrete precedents. Its [2011 chemical-compound-space program](#) helped forge an Argonne–university collaboration that produced a [machine-learning method for molecular-energy prediction](#), using DOE-supported computing. A participant in its 2023 quantum-computing program [identified a disconnect](#) between near-term and fault-tolerant research communities, leading to a [2026 bridging workshop](#). Multiple IPAM programs involved scientists from DOE labs, most recently during the [2026 Multi-Fidelity Methods for Fusion Energy program](#). These examples illustrate how sustained interaction can generate both research results and continuing collaboration.

[ICARM, the Institute for Computer-Aided Reasoning in Mathematics](#), has a complementary mission: helping mathematicians use emerging reasoning technologies, bringing mathematical and computing communities together, and broadening access while keeping mathematics central. Its scope includes formalization, automated reasoning, and machine learning. The joint

[“Lean for PDEs” workshop with SLMATH](#) offers a directly relevant example of connecting subject specialists with formalization expertise.

Closer collaboration between DOE scientists and the other communities represented could make these questions concrete. [Sottile](#)’s multiphysics perspective at Lawrence Livermore and [Samuel Pollard](#)’s work on high-consequence verification at Sandia suggest opportunities to connect laboratory applications with PDE analysis, numerical analysis, proof-assistant development, and scientific machine learning. Climate-model developers contribute complementary experience in statistical testing and diagnosis. Joint studies of selected simulation components could connect mathematical assumptions to executable behavior, while laboratory constraints help identify useful guarantees. Institute residencies and continuing working groups could sustain this two-way exchange, with publicly shareable examples providing common ground for research and training.

The workshop also suggested a concrete basis for US–Japan collaboration. [Tanaka](#)’s certified enclosures, [Yaguchi](#)’s structure-preserving learning, and [Hashimoto](#)’s interpretable model discovery could connect with the formalization and laboratory expertise represented. A proposal raised in the final panel for [jointly developed benchmarks, datasets, and evaluation criteria](#) offers a practical starting point: shared PDE and inverse-problem examples could compare accuracy, stability, certification cost, and assumptions across methods. Reciprocal visits, joint trainee projects, and independent cross-checking could turn these scientific connections into lasting collaborations.

The [US–Japan AI-for-science partnership announced in June 2026](#) provides a broader context for such exchanges. On the Japanese side, JST’s AI to Redesign Scientific Exploration (ARISE) is a flagship program implementing MEXT’s Basic Strategic Policy for Promoting AI for Science, aiming to transform scientific research through the application of AI across a wide range of disciplines, including mathematical science. Within the U.S.–Japan Genesis Mission, [RIKEN](#), Japan’s largest national research institute for basic science, has been designated to coordinate [Japan’s mathematical-sciences contribution](#), providing a natural bridge to domain sciences and large-scale computing. This discussion does not imply a workshop-specific agreement or assured access to resources.

6. A framework for coordinated progress

The discussions suggest a productive interplay among shared foundations, focused scientific demonstrations, and sustained communities of practice. Examples such as a compositional verified PDE solver, a certified learned solver, or assurance of an evolving multiphysics application could expose common obstacles while contributing reusable methods. The appropriate balance between proof, testing, and empirical validation will depend on each problem’s consequences, failure modes, and available theory.

Useful signs of progress would include clearer specifications, reproducible results, guarantees that connect across computational stages, independent reuse, and successful revalidation after changes in software or hardware. Verification cost, scientific insight, and practical adoption provide complementary perspectives on success. Space for exploratory mathematics and new learning methods remains important alongside demonstrations with well-defined objectives.

These directions invite dialogue across disciplines, institutions, and national boundaries. The common aim is trustworthy scientific knowledge: explicit assumptions, understood limitations, and reusable ideas. Sustained relationships can help turn the workshop's complementary perspectives into shared understanding and further discovery.

Possible future actions include promoting residential research programs with questions driven by researchers from academia, national labs, and industry. There are multiple precedents for these programs, including long programs at NSF Mathematics Research Institutes, as well as student-oriented industry-driven partnerships such as [the Research in Industrial Projects for Students \(RIPS\)](#) at IPAM, UCLA, and [G-RIPS Sendai](#) (G = Graduate-level) at MathCCS, Tohoku University. SLMath has a long standing tradition of collaboration with Japan through its summer graduate schools (most recent <https://www.slmath.org/summer-schools/1108>). Starting in spring 2027 SLMath and OIST will run coordinated month-long programs Singularities & PDEs representing natural phenomena at [SLMath in 2027](#) and at [OIST in 2028](#). These programs make progress on specific challenges by bringing together teams of researchers (representing different techniques and disciplines), focusing them on specific problems, and stimulating collaboration and development of new ideas and approaches.

Beyond accelerating existing mathematical research, the Genesis Mission could have an important unlocking effect for the mathematics community, creating new opportunities for mathematicians to engage in collaborative exploration with AI and with scientists and engineers across disciplinary boundaries. At the same time, it offers an opportunity to train a new generation of mathematicians who can use emerging AI tools to work more easily across fields, bringing mathematical thinking to a broader range of scientific and engineering challenges.