RelSciFM 2026: The workshop on Reliable Scientific Foundation Models: Design, Training, Grounding, and Verification

Yue Huang*
University of Notre Dame
Notre Dame, IN, USA
yhuang37@nd.edu

Xiaoda Wang
Emory University
Atlanta, GA, USA
xiaoda.wang@emory.edu

Yuchen Ma
Ludwig-Maximilians-Universität München
Munich, Bavaria, Germany
yuchen.ma@lmu.de

Wei Wang
University of California, Los Angeles
Los Angeles, CA, USA
weiwang@cs.ucla.edu

Carl Yang
Emory University
Atlanta, GA, USA
j.carlyang@emory.edu

Xiangliang Zhang
University of Notre Dame
Notre Dame, IN, USA
xzhang33@nd.edu

Abstract

Scientific Foundation Models (SciFMs) are emerging as a general-purpose substrate for scientific computing, integrating heterogeneous scientific data and interacting with computational tools to support generation, reasoning, and decision-making. However, their real-world scientific impact is increasingly limited by reliability: outputs can be plausible yet incorrect, weakly grounded in evidence, inconsistent with domain constraints, and difficult to reproduce or audit. RelSciFM 2026 provides a focused forum to advance Reliable Scientific Foundation Models through four tightly connected pillars: Design, Training, Grounding, and Verification. We invite contributions spanning architectures and inductive biases for scientific structure, data curation and scalable learning, retrieval- and tool-grounded modeling with traceable provenance, and rigorous verification protocols including calibration, constraint checking, stress testing, and reproducibility-oriented benchmarks. The workshop aims to bring together AI researchers and domain scientists to distill shared challenges, surface best practices, and produce community artifacts such as benchmark specifications and evaluation checklists for trustworthy SciFMs.

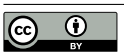
CCS Concepts

• **Computing methodologies** → **Machine learning**; *Artificial intelligence*; • **Information systems** → *Information retrieval.*

Keywords

Scientific Foundation Models, Reliability, Grounding, Verification, Trustworthy AI, AI for Science

*Corresponding authors: yhuang37@nd.edu, xzhang33@nd.edu.



KDD '26, Jeju Island, Republic of Korea

ACM ISBN 979-8-4007-2259-2/2026/08
https://doi.org/10.1145/3770855.3818241

ACM Reference Format:

Yue Huang, Xiaoda Wang, Yuchen Ma, Wei Wang, Carl Yang, and Xiangliang Zhang. 2026. RelSciFM 2026: The workshop on Reliable Scientific Foundation Models: Design, Training, Grounding, and Verification. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '26), August 09–13, 2026, Jeju Island, Republic of Korea.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3770855.3818241

1 Introduction

Generative AI is rapidly evolving from a supportive tool into a core collaborator in scientific research, reshaping how hypotheses are proposed, models are built, and evidence is synthesized from complex data. At the center of this shift are Scientific Foundation Models (SciFMs): large-scale, cross-domain models trained on heterogeneous scientific data and adapted to a wide range of downstream tasks. SciFMs promise a unifying layer for scientific workflows—supporting knowledge synthesis and reasoning, controllable generation and design, and the integration of data-driven learning with mechanistic modeling and computational tools [2, 3].

Despite fast progress, the scientific impact of SciFMs is increasingly constrained by reliability. Scientific workloads require outputs that are not only plausible, but also grounded in evidence, consistent with domain constraints, and reproducible under clear protocols. These failures reflect end-to-end issues across the pipeline, including imperfect data and supervision, spurious correlations, weak uncertainty estimates, and evaluation practices that reward surface similarity rather than scientific correctness [1, 4–6].

RelSciFM 2026 focuses on building Reliable Scientific Foundation Models through four tightly connected pillars: **Design**, **Training**, **Grounding**, and **Verification**. **Design** examines architectures and inductive biases for scientific structure, multimodal fusion, and interfaces to tools such as simulators and solvers.

Paper Submission. We invite contributions from all relevant communities spanning machine learning and data mining, as well as scientific and engineering domains (e.g., chemistry, biology, materials science, climate and environmental science, healthcare, physics,

and responsible AI). The workshop features three tracks: **Main Paper Track** (8 pages), **Position Paper Track** (4 pages), and **Short Paper Track** (2–4 pages). Topics include, but are not limited to: (1) architecture design with scientific priors, including inductive biases for invariances, conservation laws, units, and geometry; (2) evidence grounding and provenance through retrieval, attribution, and version-aware evidence tracking; (3) tool- and simulator-integrated modeling with reliable interfaces to solvers, simulators, and symbolic engines; (4) reliable data and pretraining, including curation, contamination checks, and robust objectives; (5) uncertainty and risk control, including calibration, abstention, and risk–coverage trade-offs; (6) validity, robustness, and stress testing under constraints, counterfactuals, rare regimes, and distribution shifts; (7) reproducibility and auditing with versioned pipelines and tool-trace logs; and (8) benchmarks and community testbeds for correctness, faithfulness, constraints, calibration, and shared evaluation.

2 Schedule & Attendance & Review

Tentative Schedule. The format of the workshop is a half-day event (4 hours, 8:00 am–12:00 pm). We will invite 4 keynote speakers who are leading experts from academia and industry. The potential speakers include Prof. Jian Tang (Mila/HEC Montreal), Dr. Alix Schmidt (Dow Chemical), Prof. Xin Gao (KAUST), and Prof. Xiao Hu (Emory University). We will feature contributed papers as both oral/spotlight and poster presentations for accepted submissions. We will also highlight outstanding papers in an award ceremony to facilitate focused discussion and identify community priorities.

Time	Event
8:00 am – 8:10 am	Opening Ceremony
8:10 am – 8:45 am	Keynote Talk 1 (Invited Speaker)
8:45 am – 9:20 am	Keynote Talk 2 (Invited Speaker)
9:20 am – 9:55 am	Keynote Talk 3 (Invited Speaker)
9:55 am – 10:35 am	Coffee Break and Poster Session
10:35 am – 11:10 am	Keynote Talk 4 (Invited Speaker)
11:10 am – 11:40 am	Best Paper Spotlight
11:40 am – 12:00 pm	Award Ceremony and Closing Session

Table 1: Workshop schedule.

Important KDD 2026 Workshop Deadlines. We will follow the KDD 2026 workshop timeline: CFP by Apr. 9, submission by May 20, notification by Jun. 4, statistics to chairs by Jun. 11, and final program/materials by Jun. 22, 2026 (AoE).

Target Audience and Expected Participation. The workshop targets researchers and practitioners in data mining and machine learning (foundation models, evaluation, robustness, data-centric learning), AI-for-Science (chemistry, biology, materials, and scientific automation), as well as domain experts and industry R&D teams who deploy AI systems in scientific pipelines. We also welcome participants interested in safety, security, and governance of AI systems used in scientific environments. We anticipate approximately 80–120 attendees and 70–100 submissions, with 25–45 accepted papers primarily presented as posters (plus a small number of oral/spotlight talks).

Review Process The review process for contributed material will be double-blind to reduce institutional and author bias. We will form a wide and diverse program committee and ensure broad topical coverage while preserving merit awarded by the double-blind reviewing process. Each submission will receive at least three reviews. We will provide clear review guidelines, manage conflicts of interest by collecting COI declarations from reviewers and auditing assignments, and hold a calibration session to promote fair and consistent evaluation. Decisions will be made transparently by the organizers based on review quality and discussion. To discourage already finalized work and better support workshop-style interaction, we will explicitly evaluate novelty and, importantly, the scope for discussion. We will also mentor junior reviewers by pairing them with experienced PC members.

Pre-Workshop Outreach and Community Engagement. We will promote the CFP through SIGKDD and topical mailing lists, social media, professional networks, and the workshop website. We will also collect community input before the workshop to identify open challenges and shape the discussion agenda.

Relevance to SIGKDD. The workshop aligns closely with SIGKDD themes, including large-scale data curation, representation learning, retrieval and grounding, robustness, and rigorous evaluation. It will connect the KDD community with domain scientists to advance reliable knowledge discovery with scientific foundation models.

Rationale. As SciFMs move into tool- and simulator-integrated scientific workflows, reliability issues such as ungrounded outputs, constraint violations, miscalibrated uncertainty, and brittleness under shift have become urgent. KDD 2026 is a timely venue to establish shared benchmarks, checklists, and reproducible protocols.

Invited Speakers and Panelists. Confirmed speakers include Chang Xu (Microsoft Research), Alix Schmidt (Dow Chemical), Hangwei Qian (A*STAR), James Zou (Stanford University), Jian Tang (Mila/HEC Montréal), and Xin Gao (KAUST), covering reliable AI, AI for Science, scientific foundation models, materials informatics, bioinformatics, and molecule/protein design.

References

[1] Yue Huang, Lichao Sun, Haoran Wang, Siyuan Wu, Qihui Zhang, Yuan Li, Chujie Gao, Yixin Huang, Wenhan Lyu, Yixuan Zhang, et al. 2024. TrustLLM: Trustworthiness in large language models. *arXiv preprint arXiv:2401.05561* (2024).

[2] Michael Moor, Oishi Banerjee, Zahra Shakeri Hossein Abad, Harlan M Krumholz, Jure Leskovec, Eric J Topol, and Pranav Rajpurkar. 2023. Foundation models for generalist medical artificial intelligence. *Nature* 616, 7956 (2023), 259–265.

[3] Arun James Thirunavukarasu, Darren Shu Jeng Ting, Kabilan Elangovan, Laura Gutierrez, Ting Fang Tan, and Daniel Shu Wei Ting. 2023. Large language models in medicine. *Nature medicine* 29, 8 (2023), 1930–1940.

[4] Xiaoda Wang, Ching Chang, Defu Cao, Kaiqiao Han, Fang Sun, Yue Huang, Minxiao Wang, Chang Xu, Xiao Luo, Runze Yan, et al. 2026. Position: Beyond Prediction: Toward Verifiable Physiological Waveform Reasoning with Foundation Models and Agentic LLMs. In *The Forty-Third International Conference on Machine Learning*.

[5] Xiaoda Wang, Kaiqiao Han, Yuhao Xu, Xiao Luo, Yizhou Sun, Wei Wang, and Carl Yang. 2026. SE-Diff: Simulator and Experience Enhanced Diffusion Model for Comprehensive ECG Generation. In *The Fourteenth International Conference on Learning Representations*.

[6] Yujun Zhou, Jingdong Yang, Kehan Guo, Pin-Yu Chen, Tian Gao, Werner Geyer, Nuno Moniz, Nitesh V Chawla, and Xiangliang Zhang. 2025. LabSafety Bench: Benchmarking LLMs on Safety Issues in Scientific Labs. *To appear in Nature Machine Intelligence* (2025).