James Roggeveen, Ph.D.

j.v.roggeveen@gmail.com | (508) 280-2536 | jamesroggeveen.com | LinkedIn | Google Scholar

Agent Evaluation | Scientific Reasoning | Domain Scaling

SUMMARY

Researcher and applied mathematician building agent evaluations in scientific and engineering domains. Develops expert-designed tasks, programmatic graders, and evaluation infrastructure. Research background in differentiable scientific computing, inverse problems, and physical modeling.

RESEARCH EXPERIENCE

Google DeepMind Cambridge, MA | 2025–Present

Visiting Faculty Researcher

- Developed scalable domain-specific engineering tasks and associated programmatic graders; analyzed agent trajectories to identify competency failures and guide task and grader revisions.
- Reviewed agentic tasks and environments for scientific and engineering domains in collaboration with domain experts, assessing their realism, difficulty, and validity as capability evaluations.
- Publicly credited by Google DeepMind for foundational support to the Gemini Deep Think team.

Harvard University Cambridge, MA | 2024–Present

Postdoctoral Fellow in Applied Mathematics | Advisor: Michael P. Brenner

- Built a reusable, end-to-end pipeline for model execution, collaborative testing, programmatic grading, and symbolic verification, allowing experts to focus on developing and refining domain-specific problems.
- Co-led **HARDMath2**, a 211-problem NeurIPS 2025 benchmark, and built its collaborative authoring and testing infrastructure, enabling 40+ graduate students to develop and peer-review original mathematics problems.
- Owned the evaluation infrastructure for **CMT-Benchmark**, enabling leading condensed-matter researchers to collaboratively collect, test, refine, and evaluate 50 original research-level problems across 17 frontier language models.
- Invented a meshless, differentiable spectral basis method for PDE inverse problems on irregular geometries; require coding agents to pass independent numerical checks before reporting results, helping catch invalid implementations while extending the solver to new applications.
- Designed a differentiable framework for identifying tensor-valued constitutive laws from sparse, partial flow measurements; manuscript under review at *Nature Communications*.

Princeton University Princeton, NJ | 2019–2024

Doctoral Researcher, Mechanical and Aerospace Engineering | Advisor: Howard A. Stone

- Derived models of complex fluid systems by applying mathematical theory, numerical simulation, and experiments, resulting in four first-author publications in fluid mechanics, soft matter, and biophysics.
- Created a statistical-inference pipeline for segmenting experimental data, validating model assumptions, and extracting material properties from noisy time series, reducing a roughly 30-minute analysis to seconds.

EDUCATION

Princeton University 2024

Ph.D., Mechanical Engineering

University of Cambridge 2019

Master of Advanced Study, Applied Mathematics | Distinction

Massachusetts Institute of Technology 2018

B.S., Mechanical Engineering | GPA: 5.0/5.0

SELECTED PUBLICATIONS AND BENCHMARKS

J. V. Roggeveen*, E. Y. Wang*, *et al.* "HARDMath2: A Benchmark for Applied Mathematics Built by Students as Part of a Graduate Class." *NeurIPS Datasets & Benchmarks*, 2025. arXiv:2505.11774. **Equal contribution.*

H. Pan, **J. V. Roggeveen**, *et al.* "CMT-Benchmark: A Benchmark for Condensed Matter Theory Built by Expert Researchers." *ICLR*, 2026. arXiv:2510.05228.

A. M. Sunol*, **J. V. Roggeveen***, M. G. Alhashim, H. S. Bae, and M. P. Brenner. "Learning Constitutive Models and Rheology from Partial Flow Measurements Using Automatic Differentiation." Under review at *Nature Communications*. arXiv:2510.24673.

J. V. Roggeveen and M. P. Brenner. "Meshless Solutions of PDE Inverse Problems on Irregular Geometries." arXiv:2510.25752, 2025.

J. V. Roggeveen, H. Wang, Z. Shi, and H. A. Stone. "A Calibration-Free Model of Micropipette Aspiration for Measuring Properties of Protein Condensates." *Biophysical Journal*, 2023. DOI:10.1016/j.bpj.2023.09.018.

TECHNICAL EXPERTISE

Agent evaluation: Agentic task and environment design, expert-driven benchmark development, programmatic grading, symbolic and numerical verification, trajectory analysis, collaborative evaluation workflows

Scientific computing: Automatic differentiation, inverse problems, numerical PDEs, optimization, dynamical systems, statistical inference from experimental data, mathematical modeling

Programming: Python, JAX, SymPy, Mathematica, Git