

# AI METHODS FOR SCIENCE

CDS DS 595 · Spring 2026 · Boston University

|                    |  |                  |                      |
|--------------------|--|------------------|----------------------|
| <b>Instructor:</b> | Siddharth Mishra-Sharma  | <b>Lecture:</b>  | Mon/Wed 12:20–1:35pm |
| <b>Email:</b>      | <a href="mailto:smishras@bu.edu">smishras@bu.edu</a>               | <b>Location:</b> | CAS 218              |
| <b>TF:</b>         | Wanli Cheng ( <a href="mailto:cw11997@bu.edu">cw11997@bu.edu</a> ) | <b>Credits:</b>  | 4                    |

*Last updated: January 20, 2026*

**Discussion Section:** Tuesdays 11:15am–12:05pm, MUG 205

**Office Hours:** Tue 3–4pm or by appointment, CDS 1528

## Resources:

- Course website: <https://smsharma.io/teaching/ds595-ai4science.html>
- Discussion: [Ed Discussion](#)
- Assignment/lab submission: GitHub Classroom
- Computing: GPU access through the [Shared Computing Cluster \(SCC\)](#) and LLM API/finetuning credits will be provided for project work

There is no required textbook. Optional readings will be drawn from research papers and online resources.

**Course Description:** AI is increasingly central to how science gets done, spanning simulation, experiment, theory, and observation. This course aims to equip students with the methods to understand and carry out research at the intersection of AI and the natural sciences. Topics include probabilistic inference, neural networks that encode physical symmetries and domain knowledge, generative models for scientific data, and simulation-based inference. *While framed in terms of scientific applications, the methods discussed extend well beyond scientific research, with broad applicability across industry and general AI R&D.*

Applications are drawn from domains including physics, materials science, and biology. The course involves two assignments emphasizing method design and critical analysis in collaboration with AI tools, plus two projects: a midterm applying AI methods to a scientific problem, and a final project finetuning an LLM to elicit a scientific capability.

A major focus of the course is on large language models and their emerging role in science. As LLMs become more capable of scientific reasoning and operating autonomously, understanding how to evaluate, adapt, and collaborate with these systems is becoming essential. We explore what it means to work alongside AI scientists, and how to critically assess their capabilities as well as limitations.

**Learning Objectives:** By the end of this course, students will be able to:

- Apply probabilistic inference and sampling methods (e.g., MCMC) to scientific problems
- Design neural networks that encode scientific domain knowledge

- Train generative models (e.g., diffusion models) to emulate scientific data distributions
- Use simulation-based inference to connect simulators with observations
- Evaluate the edges and limitations of LLM capabilities for scientific reasoning
- Develop intuitions for how to collaborate effectively with AI systems on research tasks
- Read and understand AI-for-science research papers

### Assessment:

|                                   |     |
|-----------------------------------|-----|
| Assignment 1: Inference .....     | 15% |
| Assignment 2: Architectures ..... | 15% |
| Discussion Labs .....             | 10% |
| Midterm Project .....             | 25% |
| Final Project .....               | 35% |

### Assignments & Components:

- **Discussion Labs.** Weekly in-class labs reinforce lecture material through hands-on programming. Students work through a notebook during discussion, exploring implementations and comparing results. Graded on participation and completion.
- **Assignments.** Two assignments ask students to propose a novel method (or modification to an existing method) and then understand/critique it through theoretical analysis and empirical evaluation. AI tools may be used freely, but the analysis and interpretation require critically engaging with what was produced.
  - Assignment 1: Design and stress-test a novel sampling or variational inference method
  - Assignment 2: Design and stress-test a novel neural network architecture or inductive bias
- **Midterm Project.** Teams of 2–3 conduct a mini research project applying methods from the first half of the course (inference, architectures, generative models) to a scientific problem. Choose from a suggested list or propose your own. Deliverable is a ~4 page workshop-style paper + code.
- **Final Project.** Teams of 2–3 identify a scientific capability that current LLMs struggle with, then finetune a language model to improve that capability. This is a two-stage project:
  - Proposal: Demonstrate that LLMs struggle at a specific scientific capability by curating a dataset or benchmark
  - Final report (~6 pages, NeurIPS format): Fine-tuned model and evaluation showing improvement, including artifacts used for fine-tuning (datasets, code, reinforcement learning environments)

### Timeline:

| Deliverable     | Out        | Due                                     |
|-----------------|------------|---|
| Assignment 1    | Wed Feb 4  | Wed Feb 18                              |
| Assignment 2    | Wed Feb 18 | Wed Mar 4                               |
| Midterm Project | Mon Mar 2  | Wed Apr 1                               |
| Final Project   | Wed Apr 1  | Proposal: Fri Apr 18; Report: Mon May 4 |

**Course Schedule:**

*This schedule is tentative and subject to change as the course progresses.*

|                                  | Date        | Topic  | Notes                  |
|----------------------------------|-------------|--|------------------------|
| <i>Week 1</i>                    |             |  |                        |
| L1                               | Wed, Jan 21 | Science in the Era of Computation                      |                        |
| <i>Week 2</i>                    |             |  |                        |
| L2                               | Mon, Jan 26 | Reasoning Under Uncertainty                            |                        |
| L3                               | Wed, Jan 28 | Framing Scientific Problems as ML Tasks                |                        |
| <i>Week 3</i>                    |             |  |                        |
| L4                               | Mon, Feb 2  | Learning by Sampling and Optimization                  |                        |
| L5                               | Wed, Feb 4  | Building Blocks of Learned Representations             | A1 out                 |
| <i>Week 4</i>                    |             |  |                        |
| L6                               | Mon, Feb 9  | Encoding Scientific Structure in Neural Networks I     |                        |
| L7                               | Wed, Feb 11 | Encoding Scientific Structure in Neural Networks II    |                        |
| <i>Week 5</i>                    |             |  |                        |
| —                                | Mon, Feb 16 | <i>No class</i>  | Presidents' Day        |
| L8                               | Tue, Feb 17 | Encoding Scientific Structure in Neural Networks III   | Substitute Monday      |
| L9                               | Wed, Feb 18 | Learning Distributions from Data I                     | A1 due; A2 out         |
| <i>Week 6</i>                    |             |  |                        |
| L10                              | Mon, Feb 23 | Learning Distributions from Data II                    |                        |
| L11                              | Wed, Feb 25 | Learning Distributions from Data III                   |                        |
| <i>Week 7</i>                    |             |  |                        |
| L12                              | Mon, Mar 2  | Guest Lecture (AI + bio)                               | Midterm out            |
| L13                              | Wed, Mar 4  | Differentiating Through Scientific Simulators          | A2 due                 |
| <i>Spring Recess: March 7–15</i> |             |  |                        |
| <i>Week 8</i>                    |             |  |                        |
| L14                              | Mon, Mar 16 | Learning Through Exploration                           |                        |
| L15                              | Wed, Mar 18 | Inverting Simulators I                                 |                        |
| <i>Week 9</i>                    |             |  |                        |
| L16                              | Mon, Mar 23 | Inverting Simulators II                                |                        |
| L17                              | Wed, Mar 25 | Discovering Equations from Data                        |                        |
| <i>Week 10</i>                   |             |  |                        |
| L18                              | Mon, Mar 30 | Guest Lecture (AI + astro)                             |                        |
| L19                              | Wed, Apr 1  | Guest Lecture (AI + particle collider physics)         | Midterm due; Final out |
| <i>Week 11</i>                   |             |  |                        |
| L20                              | Mon, Apr 6  | From Specialized to General Intelligence; Scaling      |                        |
| L21                              | Wed, Apr 8  | Quantifying and Predicting LLM Scientific Capabilities |                        |
| <i>Week 12</i>                   |             |  |                        |
| L22                              | Mon, Apr 13 | LLM Building Blocks                                    |                        |

|                | Date        | Topic   | Notes                   |
|----------------|-------------|---|-------------------------|
| L23            | Wed, Apr 15 | Teaching LLMs to Science                                      |                         |
| <i>Week 13</i> |             |   |                         |
| —              | Mon, Apr 20 | <i>No class</i>   | Patriots' Day           |
| L24            | Wed, Apr 22 | Learning Unified Representations Across Scientific Modalities | Proposal due Fri Apr 18 |
| <i>Week 14</i> |             |   |                         |
| L25            | Mon, Apr 27 | Frontiers   |                         |
| L26            | Wed, Apr 29 | Being a Human Scientist                                       |                         |
| <i>Finals</i>  |             |   |                         |
| —              | Mon, May 4  | <b>Final Project Due</b>                                      |                         |

**Discussion Sections:**

Tuesdays 11:15am–12:05pm in MUG 205.

| Week  | Date       | Topic                             | Notes                                 |
|-------|------------|-----------------------------------|---------------------------------------|
| —     | Tue Jan 20 | No discussion                     | First day of classes                  |
| 2     | Tue Jan 27 | Lab: JAX, Autodiff, GitHub        |                                       |
| 3     | Tue Feb 3  | Lab: Hamiltonian Monte Carlo      |                                       |
| 4     | Tue Feb 10 | Lab: Training Neural Networks     |                                       |
| —     | Tue Feb 17 | No discussion                     | Substitute Monday schedule            |
| 6     | Tue Feb 24 | Lab: Diffusion Models             |                                       |
| 7     | Tue Mar 3  | Midterm project intro + SCC setup |                                       |
| —     | Mar 7–15   | No discussion                     | Spring Recess                         |
| 8     | Tue Mar 17 | Lab: Differentiable Programming   |                                       |
| 9     | Tue Mar 24 | Midterm project work              | Midterm due Apr 1                     |
| 10    | Tue Mar 31 | Lab: Simulation-Based Inference   |                                       |
| 11–14 | Apr        | Final project work                | Proposal due Apr 18, Report due May 4 |

**Course Policies:**

- **Attendance.** Regular attendance in lectures is expected. Please notify the instructor of planned absences.
- **Late Work.** Late submissions are not accepted without prior arrangement. Extensions may be granted for documented emergencies.
- **Collaboration.** Discussion of concepts and approaches is encouraged. However, all submitted code and written work must be your own. When collaborating, you must acknowledge your collaborators.
- **AI Tools.** Learning to work effectively with AI is itself a course objective. Use AI tools freely to explore ideas, debug code, and deepen understanding. Focus on building genuine competence—understanding *why* something works, not just *that* it works. Disclose AI assistance in submissions, including its form and extent. See also the [CDS GAIA policy](#).

- **Recordings.** Recording of lectures requires instructor permission. Students approved for recording as an accommodation must limit use to personal study and may not share recordings.

**Academic Conduct:** All students are expected to read and abide by the [BU Academic Code of Conduct](#). Plagiarism includes copying or restating work or ideas of another person or AI software without citing the source. In computing coursework, this includes sharing code, reusing code across courses without permission, and uploading assignments to external sites. Please review the [examples of plagiarism provided by the BU Computer Science department](#). All suspected cases of plagiarism will be reported to the Academic Dean.

**Religious Observance:** Students observing a religious holiday during regularly scheduled class time are entitled to an excused absence. Please notify the instructor in advance to make arrangements for any missed work.

**Accommodations:** Boston University is committed to providing reasonable accommodations to students with documented disabilities. Students seeking accommodations should contact [Disability & Access Services](#) (25 Buick Street, Suite 300; 617-353-3658) as early as possible in the semester. A new Faculty Accommodation Letter (FAL) must be requested each semester; DAS will send this directly to instructors.