

Science of Evaluation of Artificial Intelligence for Life Sciences and Biotechnology

September 22-23, 2026

National Academy of Sciences Building

Washington, DC

AGENDA-AT-A-GLANCE

Overview

As AI systems become more capable, evaluation is no longer just a technical exercise—it is becoming a science. This conference brings together researchers and practitioners to define the emerging scientific method for AI evaluation: benchmarks, leaderboards, challenges, and the evidence needed to trust systems that reason, act, discover, and operate in high-stakes domains. This meeting will focus on AI models and emerging systems (e.g., agentic systems, self-driving labs) for responsible life sciences and biotechnology applications.

- **Benchmarks, Leaderboards, and Validity.** Developing rigorous, valid, and domain-aware benchmarks that reveal what AI systems can do, where they fail, and how their performance should be interpreted.
- **From Static Tests to Reasoning Systems.** Developing methods to evaluate agents, long-horizon tasks, tool use, symbolic reasoning, formal verification, and AI-generated knowledge that cannot be checked by fixed answer keys alone.
- **Evaluation for Science, Safety, and Society.** Applying rigorous evaluation to scientific discovery, medicine, biology, drug discovery, safety and security, and infrastructure and community consortium—where the stakes of unreliable AI are highest.

September 22, 2026

8:00am–8:45am	Registration and Breakfast
8:45am–9:15am	Welcome and Introduction
9:15am–9:45am	Call to Action: Building the Science for AI Evaluation
9:45am–10:45am	Case Studies in AI Evaluation: What Worked, What Failed, What Scaled
11:00am–11:45am	Science of Artificial Intelligence Evaluation
11:45am–12:30m	Ensuring Valid AI Evaluation in a Contaminated World
1:30pm–2:30pm	BEACON: A Benchmarking, Evaluation and Assessment Consortium for Science
3:00pm–3:45pm	The Statistics and Psychometrics of AI Evaluation
3:45pm–4:30pm	Safety, Risk, and Adversarial Evaluation
4:30pm–5:30pm	The View from Washington 1: The Landscape for Funding AI Evaluations
5:30pm–7:30pm	Reception

September 23, 2026

8:00am–9:00am	Registration and Breakfast
9:00am–9:55pm	Common Resources: Scientific Methods and Infrastructure for AI Evaluations
10:00am–10:30am	The View from Washington 2: Recent Developments in National AI Policy
10:30am–11:30am	Exploratory Discussion: Applying Science of Artificial Intelligence Evaluation to Scientific Discovery
12:30pm–2:30pm	Applying Knowledge of Benchmarks to Application-Specific Discussions - Evaluation of AI in Medicine and Healthcare - Evaluation of AI Models for Drug Discovery - Evaluation of AI for Public Health - Evaluation of AI for Biological Discovery and Engineering
3:00pm–4:00pm	Evaluating Beyond the Frontier: Agents, Unknown Knowledge, and Long-Horizon Tasks
4:00pm–4:30pm	The View from Washington 3: Recent Developments in AI Regulations and Safety
4:30pm–5:00pm	Keynote Address: A Path Forward

All times are Eastern Time.