Machine Learning for Science: Bridging Data-driven and Mechanistic Modelling Approaches

Preparatory activity

To help us understand each other's perspectives from the start of the session, we're inviting participants to send us a short (one or two sentences) response to the following question before the seminar: **What development in AI over the last five years do you think will have most impact in science over the next five years?**

We'll make a summary of responses to display and add to during the meeting.

During the meeting

We hope this workshop will provide an opportunity to explore the frontiers of AI for science, its current applications, and emerging research directions. The organisers hope to work with participants to shape each day's discussions in line with the group's interests. This agenda sets out a loose schedule of talks and group discussions.

During the meeting, if there is an issue or topic you'd like to explore in more depth, please do feel free to suggest this. We'd particularly welcome 'reflections' from the group during each session – if a talk or discussion sparks ideas, and you'd like to volunteer to give a reflection on that subject, we'd be happy to accommodate this.

Agenda

Session 1 (Monday): Learning from recent experiences

This opening session will share insights from recent work deploying AI for scientific discovery. It will:

1. Explore how AI is currently being used to enhance research and innovation.
2. Consider what lessons from these experiences are relevant to the wider application of AI for scientific discovery.
3. Identify domains or topics where participants believe AI could play an important role in unlocking further progress.

While timings will flex according to the group's focus, the following gives an overview of the day's running order:

09:00	Welcome and introductions
Talks (and breaks)	09:15 – 12:30 (to include morning coffee) • Markus Reichstein: Machine-learning-model-data-integration for a better understanding of the Earth System • Dina Machuve: Poultry Diseases Diagnostics Models using Deep Learning • Siddharth Sharma-Mishra: Finding new physics Lunch (12:30 – 13:30) 13:30 – 18:00 (including an afternoon break) • Maren Buttner: Single-cell transcriptomics • Christian Igel: Estimating ecosystem properties • Ieva Kazlauskaite: Partial differential equations and Variational Bayes • Francisco Vargas: The Schrödinger bridge problem

Group discussion and plenary feedback	16:30 – 17:00 Table discussions will consider: Where are there further opportunities to advance science through the application of AI? (What did we miss today?) What lessons or insights from today's talks could support the wider use of AI? 17:00 – 17:30 Plenary feedback and discussion

Session 2 (Tuesday): Building effective simulations

This thematic session will:

1. Review recent progress in the use of simulation in the context of AI for science and the opportunities for further progress in this field;
2. Identify current challenges or issues in the field;
3. Develop ideas for how to overcome them, whether through technical AI advances, new approaches to deployment, or other enablers.

While exact timings will flex according to the group's focus, the following gives an overview of the day's running order:

9:00	Welcome and introductions
Talks	09:00 – 10:00: Plenary discussion: Table discussions will consider: Where are there further opportunities to advance science through the application of AI? (What did we miss today?) What lessons or insights from today's talks could support the wider use of AI? 10:00 – 12:15 (to include a coffee break) • Philipp Hennig: Simulation and scientific computing • Hans Kersting: ODE filters and smoothers: probabilistic numerics for mechanistic modelling • Jakob Macke: Simulation based inference for scientific discovery: Opportunities & challenges 12:15 – 13:30 (lunch) 13:30 – 17:00 (to include an afternoon break) • Jakob Macke: Simulation based inference for scientific discovery: Opportunities & challenges • Gilles Louppe: Towards reliable simulation-based inference • Tom Dietterich: Modelling the data collection process
Group discussion and plenary feedback	17:00 – 18:00 Table discussions and plenary feedback will consider: What change would drive a step-change in our ability to make and use simulations for science? Who can drive that change (and how)?

Session 3 (Wednesday): Connecting data to causality

This thematic session will:

1. Articulate the current limitations of AI for science methods with respect to causality, and the reasons for those limitations.
2. Consider what further progress could be unlocked, if current limitations were addressed.
3. Identify how technical advances could help overcome these limitations.

While exact timings will flex according to the group's focus, the following gives an overview of the day's running order:

9:00	Welcome
Talks	09:00 – 10:00 Plenary discussion: What are the fundamental 'ML Science' questions? 10:00 – 12:15 (including morning coffee break) • Bernhard Schoelkopf: Causality, causal digital twins, and their applications • Jonas Peters: Invariance: From Causality to Distribution Generalization • Niki Kilbertus: Can we discover dynamical laws from observation? 12:15 – 13:15 (lunch) 13:15 – 15:30 (walk) 15:30 – 15:45 (coffee) 16:00 – 18:00 • Soledad Villar: Invariances and equivariances in machine learning • Bubacarr Bah: Divide-and-Conquer Equation Learning with R2 and Bayesian Model Evidence

Session 4 (Thursday): Encoding domain knowledge

This session will:
1. Identify what strategies or techniques can connect domain knowledge to the design of ML systems;
2. Develop ideas for how to deepen these connections, whether through technical or operational interventions.

9:00	Welcome
	09:00 – 09:45 Plenary feedback: What change would drive a step-change in our ability to make causal connections using AI? Who can drive that change? 09:45 – 12:15 (including coffee break) • Sami Kaski: Virtual laboratories for science, assisted by collaborative AI • David Hogg: Integrating the laws of physics in ML 12:15 – 13:15 (lunch) 13:15 – 15:30

	• Mauricio Alvarez: Latent force models • Carl Henrik Ek: Translating mechanistic understandings to stochastic models (coffee)
Group discussion and plenary feedback	15:45 – 16:30 Session plenary: How can we deepen the connections between AI and domain knowledge? What technical interventions can help? What forms of collaboration between ML and domain experts are useful? 16:30 – 18:00 Workshop plenary: A reflective session will consider the messages emerging from the week's discussions. The week in review: what have we learned this week about the current state of 'AI for science'? *Participants will consider highlights or key insights from the meeting, and their implications for the field.* Looking ahead: what action or change is needed over the next 5 years to drive further use of AI for research and innovation? Who can help drive that change? *Participants will co-design a roadmap for the development of AI for science.*