Data-driven discovery in the chemical sciences

10-12 September 2024 | Oxford, UK

Faraday Discussions

Tuesday 10 September

11:00	Registration and refreshments
12:00	Lunch
12:45	**Welcome and introductions** Volker Deringer & Fernanda Duarte, Co-c*hairs of Scientific Committee*
12:55	**Outline of Discussion format** Emma Gorrell & Lauren Yarrow-Wright, *Royal Society of Chemistry*
13:00	**Introductory Lecture – Spiers Memorial Lecture** (Session chair: Volker Deringer) Alán Aspuru-Guzik *University of Toronto, Canada*
14:00	Short break (no refreshments)
	Session 1: Discovering chemical structure (Session chair: Graeme Day)
14:15	**Fast and easy structure prediction enables discovery** Chris Pickard *University of Cambridge, UK*
14:20	**Integration of generative machine learning with the heuristic crystal structure prediction code FUSE** Chris Collins *University of Liverpool, UK*
14:25	**Large property models: A new generative paradigm** Brett Savoie *Purdue University, USA*
14:30	Discussion
15:45	Refreshments
	Session 1: Discovering chemical structure (Session chair: tbc)
16:15	**Fine-tuning foundational machine learning potentials for bespoke systems** Venkat Kapil *University of Cambridge, UK*
16:20	**Knowledge distillation of neural network potential for molecular crystals** Takuya Taniguchi *Waseda University, Japan*
16:25	**Modelling ligand exchange in metal complexes with MLPs** Veronika Juraskova *University of Oxford, UK*
16:30	Discussion
17:45	Flash poster presentations (by invitation of the Scientific Committee)
18:00	Poster session and wine reception
19:30	Close

Wednesday 11 September

	Session 2: Discovering structure–property correlations (Session chair: tbc)
09:00	**Facilitating guided materials discovery with data-driven optimization** Kim Jelfs *Imperial College London, UK*
09:05	**Sequence determinants of protein phase separation and recognition by protein phase-separated condensates through molecular dynamics and active learning** Arya Changiarath Sivadasan *Johannes Gutenberg University of Mainz, Germany*
09:10	**Discovering highly anisotropic dielectric crystals** Yuchen Lou, Alex Ganose *Imperial College London, UK*
09:15	Discussion
10:30	Refreshments
	Session 2 continued: Discovering structure–property correlations (Session chair: tbc)
11:00	**Leveraging natural language processing to curate the tmCAT, tmPHOTO, tmBIO, and tmSCO datasets of functional transition metal complexes** Heather Kulik *Massachusetts Institute of Technology, USA*
11:05	**Are we fitting data or noise? Analysing the predictive power of commonly used datasets in drug-, materials-, and molecular-discovery** Daniel Crusius *University of Oxford, UK*
11:10	**Using prediction rigidities to assess learning dynamics and guide dataset construction in data-driven chemistry** Sanggyu Chong *EPFL, Switzerland*
11:15	**Active thermochemical tables: Accurate and reliable thermochemistry by data analysis of complex thermochemical networks** Branko Ruscic *Argonne National Laboratory, USA*
11:20	Discussion
13:00	Lunch
	Session 3: Discovering trends in big data (Session chair: tbc)
14:00	**Paper title tbc** Jean-Louis Reymond *University of Bern, Switzerland*
14:05	**Specialising and analysing instruction-tuned and byte-level language models for organic reaction prediction** Jiayun Pang *University of Greenwich, UK*
14:10	**What can we learn from 1,000 crystal structure predictions? Building on experimental data with large-scale computation** Christopher Taylor *University of Southampton, UK*
14:15	Discussion
15:30	Refreshments
	Session 3 continued: Discovering trends in big data (Session chair: tbc)
16.00	**Paper title tbc** Claudia Draxl *Humboldt-Universität zu Berlin, Germany*
16:05	**Next generation of the InChI as molecular identifier** Gerd Blanke, Sonja Herres-Pawlis

	RWTH Aachen University, Germany
16:10	**Optical materials discovery and design via federated databases and machine learning** Matthew Evans *UCLouvain, Belgium*
16:15	Discussion
17:30	Close of sessions
19:00	Pre-dinner drinks
19:30	Conference dinner

Thursday 12 September

	Session 4: Discovering synthesis targets (Session chair: tbc)
09:00	**Analysis of uncertainty of neural fingerprint-based models** Miriam Mathea *BASF, Germany*
09:05	**Re-evaluating retrosynthesis algorithms with syntheseus** Marwin Segler *Microsoft Research AI for Science, UK*
09:10	**Embedding human knowledge in uncertainty-aware synthesis rules** Basita Das *Massachusetts Institute of Technology, USA*
09:15	Discussion
10:30	Refreshments
	Session 4 continued: Discovering synthesis targets (Session chair: Fernanda Duarte)
11:00	**Mapping inorganic crystal chemical space** Aron Walsh *Imperial College London, UK*
11:05	**A critical reflection on attempts to machine-learn materials synthesis insights from text-mined literature recipes** Wenhao Sun *University of Michigan, Ann Arbor, USA*
11:10	Discussion
12:00	**Concluding remarks lecture** (Session chair: Fernanda Duarte) Andy Cooper *University of Liverpool, UK*
12:30	**Acknowledgements**
12:45	Close of meeting and lunch

Please note that this is a draft programme and timings may change.