nature conferences

AI for Discovery and Research Automation

Virtual Conference | October 16 – 17, 2025

Organized by:

nature communications

nature machine intelligence

AI for Discovery and Research Automation

October 16 – 17, 2025

VIRTUAL CONFERENCE

https://conferences.nature.com/

Organizers:

Anita Chandran
Nature Communications

William Henson
Nature Machine Intelligence

Manel Mondelo
Nature Communications

Steven Lukman
Nature Communications

Table of Contents

GENERAL INFORMATION

The conference recording will be available ~48 hours after the conference conclusion for registrants to view for up to 60 days.

Scientific Session Protocol

Audio and/or video recording of the scientific sessions is not permitted.

Abstracts may not be re-printed without permission of the presenter.

Code of Conduct

Nature Conferences is committed to promoting practices that support diversity and inclusion. We value wide-ranging views, expertise, opinions, backgrounds and experiences. We intend Nature Conferences to be welcoming, safe, collaborative and productive for all attendees, volunteers and staff in both a live and virtual setting.

If you need to report an issue of feeling unsafe or harassed, please email annie.zhou.1@nature.com immediately.

Those who are deemed to be acting inappropriately will be expelled from this conference, at the discretion of the conference organizers.

Internet Access

A strong WiFi connection is key to a successful virtual conference. Please make sure that your WiFi signal is strong enough to support streaming video. If you do happen to lose connectivity, please know that all sessions will be available for viewing ~48 hours after the conference and for the next 60 days.

Our partnered streaming platform StreamGo is most compatible with **Google Chrome**. Please ensure to download Google Chrome to your device before the start of the conference.

Conference Certificates

If you are in need of a certificate of attendance and/or short talk presentation, please email Annie Zhou at annie.zhou.1@nature.com after the conference concludes.

AI for Discovery and Research Automation

October 16 – 17, 2025

Times are displayed in British Summer Time (UTC +1)
*Talk titles followed by * are selected talks*

Day 1: October 16, 2025, 9:50 – 16:40

Session I: AI in Scientific Discovery

Chair: Anita Chandran (*Nature Communications,* UK)

9:50 – 10:00 Welcome and Opening

Anita Chandran (*Nature Communications*, UK)
William Henson (*Nature Machine Intelligence,* UK)
Manel Mondelo (*Nature Communications*, Germany)
Steven Lukman (*Nature Communications*, UK)

10:00 – 10:30 *Harnessing AI for Small Molecule Drug Discovery*

Lauren Therese May (Monash University, Australia)

10:30 – 11:00 *Trying to Make the Novel Exceptional with Generative AI for Transition Metal Chemistry*

David Balcells (Oslo University, Norway)

11:00 – 11:30 *Building LLM agents for robotic AI-chemists*

Linjiang Chen (University of Science and Technology of China, China)

11:30 – 12:00 *Reinforcement Learning for Configurational Design of High-PCE Chalcogenide Photo-Absorbers*

Zhenzhu Li (Imperial College London, UK)

12:00 – 13:00 Break

Session II: Generative AI Powered Discovery

Chair: William Henson (*Nature Machine Intelligence,* UK)

13:00 – 13:45 *Towards Deployment-Centric Generative and Multimodal AI for Discovery*

Haiping Lu (University of Sheffield, UK)

13:45 – 14:15 *Physics-Guided Enzyme Engineering: Large-Scale Kinetic Data Extraction and Multispecificity Prediction*

Zhongyue (John) Yang (Vanderbilt University, USA)

14:15 – 14:45 *Virtual Patient Labs: AI-Driven Simulation and Diagnostics for Precision Oncology*

Charlotte Bunne (EPFL, Switzerland)

14:45 – 15:15 *Vendi Scoring For Discovery*

Adji Bousso Dieng (Princeton University, USA)

15:15 – 15:25	*LLM-Assisted Ensemble Learning for High-Performance Energy Storage Transition Metal Oxide Materials**
	Xiangyu Cao (Hanyang University, Korea)
15:25 – 15:35	*Computational Data Driven Methods in Discovery of Native Neurological Transporter Modulators**
	Taner Karagöl (Istanbul University Medical Faculty, Turkey)
15:35 – 15:45	*Characterising Proteins with AI: Language Models Driving Scientific Discovery**
	Deepak Chaurasiya (Indian Institute of Information Technology Allahabad India, India)
15:45 – 15:55	*Machine Learning Approaches to Surpass the Limitations of the Beer-Lambert Law**
	Muthuchamy Murugavel (Shri Ramasamy Memorial (SRM) University Sikkim, India)
15:55 – 16:40	**Panel Discussion – AI ethics**
	Chair: Anita Chandran (*Nature Communications,* UK)
	Panelists: Trenton Jerde (*Nature Machine Intelligence,* US)
	Yann Sweeney (*Nature,* UK)
	Manel Mondelo *(Nature Communications,* Germany)

Day 2: October 17, 2022, 15:00 – 19:25

Session III: Characterisation Methods

Chair: Steven Lukman (*Nature Communications,* UK)

9:50 – 10:00	Welcome Remarks
	Anita Chandran (*Nature Communications*, UK) **William Henson** (*Nature Machine Intelligence,* UK) **Manel Mondelo** (*Nature Communications*, Germany) **Steven Lukman** (*Nature Communications*, UK)
10:00 – 11:00	**Meet-the-Editor Session**
	Chair: Anita Chandran (*Nature Communications,* UK)
	Panelists: William Henson (*Nature Machine Intelligence,* UK) **Manel Mondelo** (*Nature Communications*, Germany)
11:00 – 11:30	*From Machine Learning to Biological Insight: Explainable AI for Decoding Immune Recognition*
	Jiangning Song (Monash University, Australia)
11:30 – 12:00	*How should we use AI to help us find new antibiotics*
	Jonathan Stokes (McMaster University, Canada)
12:00 – 13:00	Break

Session IV: Self-driving Laboratories

Chair: Manel Mondelo (*Nature Communications,* Germany

13:00 – 13:45	*Building a Global Infrastructure for AI-Driven Innovation*
	Jun Jiang (University of Science and Technology of China, China)
13:45 – 14:15	*Machine Learning Hits the Lab: Practical Experiment Planning for Molecular Discovery*
	Felix Strieth-Kalthoff (University of Wuppertal, Germany)
14:15 – 14:45	*Self-driving Laboratory Empowering Innovation in Functional Polymer Discovery*
	Jie Xu (Argonne National Laboratory, USA)
14:45 – 15:15	*Advancing Pharmaceutical Discovery Through AI-Driven Experimental Design and Data Architecture Innovation*
	Melodie Christensen (Merck, USA)
15:15 – 15:25	*LLM-Guided Evolutionary Monte Carlo Tree Search for Explainable Algorithm Discovery in Gravitational-Wave Detection**
	He Wang (University of Chinese Academy of Science, China)
15:25 – 15:35	*ANN-Enabled Research Automation for Discovery and Translation: Application to Intranasal CNS Drug Delivery**
	Ujban Hussain (Rashtrasant Tukadoji Maharj Nagpur University, Nagpur, India)
15:35 – 15:45	*Reinforcement Failing guides the discovery of emergent physical dynamics in adaptive tumor therapy**
	Jona Kayser (Max Planck Zentrum für Physik und Medizin, Germany)
15:45 – 15:50	**Closing remarks**
	Anita Chandran (*Nature Communications*, UK) **William Henson** (*Nature Machine Intelligence,* UK) **Manel Mondelo** (*Nature Communications*, Germany) **Steven Lukman** (*Nature Communications*, UK)

SPEAKERS (IN PROGRAM ORGER)

* are Selected Talk (ST) presenters

Lauren May

Monash University, Australia

Associate Professor Lauren May, Monash Institute of Pharmaceutical Sciences (MIPS), is Head of the Cardiac GPCR Biology laboratory and leads the MIPS Climate Health Focus Area and MIPS Green Lab Community of Practice. Integrating structural information with pharmacology, artificial intelligence, computational biology and translational disease models, Lauren's research has advanced our understanding of G protein-coupled receptor (GPCR) allosterism and biased agonism. A/Prof May has published 81 manuscripts (e.g. Nature, Cell, PNAS, Nat Mach Intell), secured competitive research funding (>$15M last 10 years) and supervised 12 PhD completions. Leadership roles in the MIPS Climate Health focus area (lead, 2024-present), Australian Cardiovascular Alliance Drug Discovery Flagship (advisory member, 2019-present) and MIPS TPA (coordinator, 2020-2024) demonstrate her commitment to facilitating impactful cross-disciplinary research. An advocate for diversity in science, she co-founded Her Research Matters (2021 Vice-Chancellor's Award; 2024 Faculty of Pharmacy and Pharmaceutical Sciences Award).

Harnessing AI for Small Molecule Drug Discovery

The process of small molecule drug discovery remains an expensive and time-consuming process. Computational methods can predict drug-receptor interactions, accelerating the exploration of chemical space and enabling an improved rate of hit identification1. These methods predict drug-target interactions using di=erent types of input data. Sequence-based approaches rely on one-dimensional (1D) inputs, such as ligand SMILES and protein sequences, or two-dimensional (2D) representations, including molecular graphs and predicted contact maps. Structure-based approaches incorporate three-dimensional (3D) protein structures, determined experimentally or predicted in silico (e.g., via AlphaFold2).
We developed an AI-based, structure-independent framework to identify small molecule ligands targeting G protein-coupled receptors (GPCRs). Our physicochemical graph neural network, known as PSICHIC, can predict small molecule binding affinity and functional effect directly from sequence data2. By learning interpretable interaction fingerprints, PSICHIC achieves state-of-the-art accuracy in virtual screening tasks, even in the absence of structural information. In a screening campaign for A1R agonists, PSICHIC successfully ranked the sole active novel compound within the top three candidate compounds. The residue- and atomlevel interpretability provided by PSICHIC offers insights into the molecular determinants of binding and selectivity. This study supports the integration of artificial intelligence frameworks to facilitate the rapid progress of early-phase drug discovery.

1. Nguyen. A. T. N., et al. (2023) The application of artificial intelligence to accelerate G protein-coupled receptor drug discovery. Br J Pharmacol 181(14):2371-2384. doi.org/10.1111/bph.16140

2. Koh, H.Y. et al. (2024) Physicochemical graph neural network for learning protein–ligand interaction fingerprints from sequence data. Nat Mach Intell 6:673–687. doi.org/10.1038/s42256-024-00847-1

David Balcells

Oslo University, Norway

David Balcells completed his PhD studies in 2006 in the Maseras group at ICIQ, on the topic of computational asymmetric catalysis. After a postdoc with Odile Eisenstein in the University of Montpellier, he became a Juan de la Cierva fellow in the Autonomous University of Barcelona, working on C-H activation in open-shell systems. In 2012, he moved to the University of Oslo and, after an MSCA postdoctoral fellowship, he started his independent career as a Principal Investigator at the Hylleraas Research Center of Excellence, where he was promoted to Research Professor in 2022. At the Hylleraas Centre, David is leading a research group advancing the application of machine learning to transition metal chemistry. He has received several awards, including the Ground-Breaking Research Grant from the Norwegian Research Council and the Young Researcher Award from the Spanish Royal Society of Chemistry.

Trying to Make the Novel Exceptional with Generative AI for Transition Metal Chemistry

I will present the tmQM dataset series that we developed since 2020, already providing diverse data (energies, geometries, excited states, quantum properties) for thousands of transition metal complexes. These datasets were used to develop predictive models based on deep graph learning methods. More recently, we have extended this research to generative models based on variational autoencoders (VAEs) and genetic algorithms (GAs). Whereas the GAs enabled multi-objective optimization tasks within very large chemical spaces (up to billions of metal complexes), the VAEs enabled the generation of novel metal ligands and complexes with high chemical validity and novelty. Integrating these two approaches in an evolutionary machine learning model, we are now investigating the inverse design of novel metal chromophores that are optimal over three different properties: water solubility, and visible light absorption intensity and broadness.

Linjiang Chen

University of Science and Technology of China (USTC), China

Linjiang is a professor of digital chemistry at the University of Science and Technology of China (USTC). He currently serves as Deputy Chief Engineer for the Chinese Academy of Sciences's “Intelligent Scientist” program and as head of the research division of intelligent chemistry in the State Key Laboratory of Precision and Intelligent Chemistry, USTC. His research focuses on the precision, rational design and intelligent, efficient exploration of chemical and materials systems empowered by computer science and automation. He has developed a variety of interdisciplinary, cross-domain methods, technologies, and tools that integrate chemicalmaterials theory, artificial intelligence, big-data techniques, and chemical automation. To date, he has published 80+ papers in leading journals such as Nature, Science, Nature Materials, Nature Chemistry, Nature Communications, and the Journal of the American Chemical Society, with more than 8,000 citations and an h-index of 36.

Building LLM agents for robotic AI-chemists

Autonomous experimentation promises to compress the path from idea to insight in chemistry, but most demonstrations still struggle with multi-step, multi-instrument tasks and real-lab variability. In this talk, I will present a robotic AI chemist built on ChemAgents—a hierarchical, multi-agent framework running an on-board Llama-3.1-70B model—that plans, executes, and learns from complex experiments with minimal human input. A Task Manager coordinates four role-specialized agents (Literature Reader, Experiment Designer, Robot Operator, and Computation Performer) coupled to four foundational resources (a literature database, protocol library, model library, and automated lab). I will show end-to-end results across seven tasks from “make & measure” characterization (e.g., FTIR of azobenzenes; PXRD of metal oxides; color-pure perovskite QD films) and parameter screening (full-factorial synthesis of g-C_3N_4 for HER; BiOX photocatalytic degradation of tetracycline), to a discovery-and-optimization workflow that mined literature, synthesized 100 quinary metal–organic high-entropy catalysts (MO-HECs) for the oxygen evolution reaction, trained a fused model, and closed the loop with Bayesian optimization to identify and validate an optimal composition (266.1 mV at 10 mA cm^{-2}). Finally, I will highlight portability by deploying ChemAgents—without bespoke retraining—in a different robotic lab to autonomously run photocatalytic organic reactions with GC–MS monitoring. Together, these results outline a practical path toward on-demand autonomous chemical research that accelerates discovery while democratizing access to advanced experimental capabilities.

Zhenzhu Li

Imperial College London, UK

Dr. Zhenzhu Li is an AI in Science Fellow at Imperial College London. Her research lies at the intersection of computational materials science and artificial intelligence, with a focus on solar energy materials. She is particularly interested in defect physics, growth conditions, recombination-limited photovoltaic performance, and the inverse design of novel photovoltaic materials through generative AI models. Dr. Li has published extensively in leading scientific journals, including Nature Materials, Nature Communications, Advanced Functional Materials, Energy Storage Materials, and Matter. She is a recipient of the 2021 Young Researcher Award from the International Conference on Advanced Materials and Devices (ICAMD) and was awarded the prestigious Eric and Wendy Schmidt AI in Science Fellowship at Imperial College London. Looking ahead, her research aims to address the urgent need for AI-driven materials discovery by elucidating defect–structure–property relationships and advancing intelligent generative frameworks for next-generation photovoltaic materials.

Reinforcement Learning for Configurational Design of High-PCE Chalcogenide Photo-Absorbers

Zhenzhu Li[1], Aron Walsh[1]
[1]Department of Materials, Imperial College London, Exhibition Road, London SW7 2AZ, United Kingdom

The configurational design of photovoltaic materials through alloying is one of the crucial techniques for achieving their highest potentials in solar energy conversion. However, the synergy of elements within the alloyed crystal often correlates strongly with its configurational site occupancies of constituent elements. In this work, we test Bayesian optimisation (BO), genetic algorithms (GA), and reinforcement learning (RL) as strategies for optimising photovoltaic materials. We show that reinforcement learning (RL) provides a flexible, interpretable and efficient framework for such optimizations. Focusing on three types of alloyed chalcogenide photo absorbers – the chalcogenide perovskites $BaZr(S, Se)_3$, the rocksalt $AgBiS_2$ and the Kesterite $(Cu, Ag)_2ZnSnS_4$, our trained RL models employing A2C[1] and PPO[2] policy learning strategies, showed higher learning capacity and smaller performance variance compared to the BO and GA models. Besides, the interpretability of our RL models allows insights into material configurations by analysing learned action probabilities. Benefitting from the fine-tunable parameters of neural networks, we envision that the transfer of simulation-based RL models to real experiments will hold an important role in future automated laboratories.

[1] Volodymyr Mnih, Adrià Puigdomènech Badia, Mehdi Mirza, Alex Graves, Timothy P. Lillicrap, Tim Harley, David Silver, Koray Kavukcuoglu, https://doi.org/10.48550/arXiv.1602.01783
[2] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, Oleg Klimov, https://doi.org/10.48550/arXiv.1707.06347.

Haiping Lu

University of Sheffield, UK

Haiping Lu is a Professor of Machine Learning at the University of Sheffield, UK, where he leads AI Research Engineering at the Centre for Machine Intelligence. He is also Director of the UK Open Multimodal AI Network (UKOMAIN), funded by the UK's Engineering and Physical Sciences Research Council (EPSRC). His research focuses on translational multimodal AI, integrating diverse types of data to tackle challenges in healthcare and scientific discovery, with methodological interests in foundation models, generative AI, domain adaptation, and transfer learning.

His recent work spans brain and cardiac imaging, cancer diagnosis, protein design, and drug and materials discovery. He leads the development of PyKale, an open-source Python library for knowledge-aware machine learning. He serves as an Associate Editor for IEEE Transactions on Neural Networks and Learning Systems and IEEE Transactions on Cognitive and Developmental Systems, and has received awards from the Alan Turing Institute, Amazon, the Wellcome Trust, and the UK's National Institute for Health and Care Research.

Towards Deployment-Centric Generative and Multimodal AI for Discovery

Generative and multimodal artificial intelligence (AI) are opening new frontiers in scientific discovery across healthcare, science and engineering. Generative approaches promise to accelerate discovery in areas such as protein design and drug development, yet they face enduring challenges of interpretability, uncertainty and deployment. We advocate a deployment-centric workflow that incorporates deployment constraints early, complementing data-centric and model-centric strategies to reduce the risk of undeployable solutions. We also emphasise deeper multimodal integration across biological, clinical and materials data, supported by stakeholder engagement and interdisciplinary collaboration. To illustrate this agenda, I will highlight recent advances in generative protein design, emerging directions in cancer research, and opportunities for open benchmarks and shared infrastructure. By combining generative innovation with a deployment-centric perspective, our community can build AI systems that are not only powerful but also trustworthy and impactful in practice.

Zhongyue John Yang

Vanderbilt University, USA

Zhongyue John Yang is the SC Family Dean's Faculty Fellow, Assistant Professor of Chemistry, Chemical and Biomolecular Engineering at Vanderbilt University. He graduated from the inaugural Chemistry Po-Ling program at Nankai University in 2013, earned his Ph.D. in Theoretical and Computational Chemistry with Ken Houk at UCLA in 2017, and undertook postdoctoral training with Heather Kulik in the Department of Chemical Engineering at MIT from 2018 to 2020. Since fall 2020, he has started his independent research group at Vanderbilt and has published 31 peer-reviewed papers as an independent investigator.

His group seeks to redefine protein engineering by anchoring it in molecular-level physical principles (Nat. Comput. Sci. 2025). They established Mutexa, a physics-augmented AI platform for predicting and designing beneficial protein variants (J. Chem. Theory Comput. 2023). Leveraging Mutexa, they established in silico tools for predicting the outcome of enzyme-catalyzed hydrolytic kinetic resolution (Chem. Sci. 2023), modifying enzymatic specificity (Chem. Catalysis 2025), designing cold adapted bidomain amylases (Angew. Chem. Int. Ed. 2025), predicting 3D structures of lasso peptides (Nat. Commun. 2025), and so on. His research is funded by U.S. National Science Foundation, National Institute of Health, and Rosetta Commons. He is a recipient of NIH MIRA Award in 2022, Robin Hochstrasser Young Investigator Award in 2023, and ACS OpenEye Junior Faculty Award in Computational Chemistry in 2024. He is a member of the Early Career Board for the Journal of Chemical Theory and Computation by the ACS Publications.

Physics-Guided Enzyme Engineering: Large-Scale Kinetic Data Extraction and Multispecificity Prediction

My group seeks to redefine protein engineering by anchoring it in physical meaning. We are developing Mutexa, a physics-informed artificial intelligence (AI) platform for "intelligent" protein engineering, enabling researchers to identify super-mutants with non-native functional performance while uncovering the molecular insights behind unpredictable experimental outcomes [1]. Protein engineering, despite decades of progress, remains reliant on labor- and resource-intensive screening that delivers only "what you screen for" and offers limited mechanistic insight. While AI promises acceleration, I question whether AI alone can yield generalizable, trustworthy models [2].

In this talk, I will present a physics-guided data and modeling pipeline for enzyme multispecificity prediction, built on EnzyExtract: our large-scale, LLM-powered system that automatically extracts and cleans enzyme–substrate–kinetics records from full-text literature. Processing 137,892 publications, EnzyExtract assembled over 218,000 kinetic entries (kcat and Km), spanning 3,569 four-digit EC classes and uncovering 89,544 kinetic entries absent from BRENDA. After sequence and chemical normalization to UniProt and PubChem, we compiled 92,286 high-confidence, model-ready records (EnzyExtractDB). Using this dataset, we train models that predict enzymes and enzyme mutants' capability (i.e., multispecificity) of catalyzing multiple, chemically diverse substrates. The model enables rational selection of enzyme scaffolds to launch directed evolution for new-to-nature catalysis.

[1] Yang, Z. J.; Shao, Q.; Jiang, Y.; Jurich, C.; Ran, X.; Juarez, R.; Yan, B.; Stull, S.; Gollu, A.; Ding, N. "Mutexa: A Computational Ecosystem for Intelligent Protein Engineering" Journal of Chemical Theory and Computation. 2023. 19, 7459–7477.
[2] Jurich, C.; Shao, Q.; Ran, X.; Yang, Z. J. "Physics-based Modeling in the New Era of Computational Enzyme Engineering" Nature Computational Science. 2025, 5, 279–291.

Charlotte Bunne

EPFL, Switzerland

Charlotte Bunne is an assistant professor at EPFL in the School of Computer and Communication Sciences (IC) and School of Life Sciences (SV). She is part of the Swiss Institute for Experimental Cancer Research (ISREC) and the EPFL AI Center. Before, she was a PostDoc at Genentech and Stanford working with Aviv Regev and Jure Leskovec and completed a PhD in Computer Science at ETH Zurich working with Andreas Krause and Marco Cuturi. During her graduate studies, she was a visiting researcher at the Broad Institute of MIT and Harvard hosted by Anne Carpenter and Shantanu Singh and worked with Stefanie Jegelka at MIT. Charlotte has been a Fellow of the German National Academic Foundation and is a recipient of two ETH Medals.

Virtual Patient Labs: AI-Driven Simulation and Diagnostics for Precision Oncology

Advances in precision oncology require methods that not only analyze but also simulate complex patient states. In this talk, I will describe our recent work on innovating new AI architectures that combine multi-modal foundation models with generative modeling for treatment effect prediction.

We design foundation models to capture the complexity of cancer by learning across biological scales, from molecular interactions to tissue architecture. Architecturally, our approach builds on principles from multi-modal representation learning and large-scale vision models, enabling the integration of proteomics, pathology, and clinical annotations into a unified virtual patient space. Through unsupervised learning and multi-scale neural network design tailored to high-dimensional multiplexed imaging, we create representations where new biopsy samples can be consistently mapped and biomarkers can be discovered, supporting downstream analyses of molecular, morphological, and spatial complexity. On top of these representations, we extend the framework with generative modeling to move from analysis to simulation, predicting therapeutic responses, resistance dynamics, and enabling in silico exploration of treatment outcomes.

Our broader vision is to develop Virtual Patient Labs as digital counterparts of patients, where integrative models and generative simulation converge to anticipate disease trajectories and treatment responses. These environments would make in silico experimentation possible at patient scale and open new avenues for AI-guided decision support in oncology.

Adji Bousso Dieng

Princeton University, USA

Adji Bousso Dieng is an Assistant Professor of Computer Science at Princeton University where she leads Vertaix on research at the intersection of artificial intelligence and the natural sciences. She is affiliated with the Chemical and Biological Engineering Department, the Princeton Materials Institute, the Princeton Quantum Initiative, the Princeton Plasma Physics Laboratory, and the High Meadows Environmental Institute (HMEI) at Princeton. She is also the founder of the nonprofit The Africa I Know. She has been recently named a Distinguished Early Career Presenter at the MRS Meeting, One of 10 African Scholars to Watch by The Africa Report, an Outstanding Recent Alumni by Columbia University's Grad School of Arts and Sciences, an AI2050 Early Career Fellow by Schmidt Futures, and as the Annie T. Randall Innovator of 2022 for her research and advocacy by the American Statistical Association. Dieng received her Ph.D. from Columbia University. Her doctoral work received many recognitions, including a Google Ph.D. Fellowship in Machine Learning, a rising star in Machine Learning nomination by the University of Maryland, and a Savage Award from the International Society for Bayesian Analysis, for her doctoral thesis. Dieng's research has been covered in media such as the New Scientist. She hails from Kaolack, Senegal.

Vendi Scoring For Discovery

This talk will cover the concepts, tools, and methods that make up Vendi Scoring, a new research direction focused on the concept of diversity. I'll begin by introducing the Vendi Scores, a family of diversity metrics rooted in ecology and quantum mechanics, along with their extensions. Next, I'll discuss algorithms for efficiently searching large materials databases and exploring complex energy landscapes, such as those found in molecular simulations, using the Vendi Scores. Finally, I'll introduce the new concept of 'algorithmic microscopy,' which stems from Vendi Scoring, and describe the Vendiscope, the first algorithmic microscope designed to help scientists zoom in on large data collections for data-driven discovery.

Xiangyu Cao*

Hanyang University，Korea

Mr. Cao is a Ph.D. candidate in the Department of Chemical Engineering at Hanyang University and received his B.Eng. degree from South China University of Technology. His research combines machine learning and ensemble learning approaches to accelerate the design and optimization of functional materials for next-generation energy storage and conversion devices. He further focuses on the scalable synthesis of metal–organic frameworks (MOFs) and transition metal compound composites for energy-related applications, with related works published in Small and Chemical Engineering Journal.

LLM-Assisted Ensemble Learning for High-Performance Energy Storage Transition Metal Oxide Materials

Xiangyu Cao[1], Mumtaz Ali[1], Shuangquan Zhang[3], Chengjia Liu[4], Min Jae Ko[1,2]

[1]Department of Chemical Engineering, Hanyang University, Seoul, Republic of Korea;
[2]Department of Battery Engineering, Hanyang University, Seoul, Republic of Korea;
[3]School of Chemistry and Chemical Engineering, South China University of Technology, Guangzhou, Guangdong, 510640, China
[4]Department of Environmental Studies for Advanced Society, Graduate School of Environmental Studies, Tohoku University, Sendai, Japan

Transition metal oxide (TMO) composites have emerged as highly promising candidates for aqueous energy storage devices due to their abundant redox-active sites, tunable electronic structures, and structural stability. However, the vast diversity of TMOs and their possible combinations with conductive polymers, carbonaceous materials, and other nanostructures poses significant challenges for rapid screening and optimization. In this study, we established a comprehensive energy-storage dataset of TMO composites with the assistance of large language models (LLMs) for automated literature data curation. An ensemble learning framework was then employed to predict device performance, achieving correlation coefficients above 0.9, outperforming individual models such as random forest (RF), multilayer perceptron (MLP), and k-nearest neighbors (KNN). Feature-importance analysis revealed that redox-active metal centers, nanoscale morphology, and operating conditions are the most critical factors governing the electrochemical performance. Guided by these insights, we designed and synthesized a novel 2D Co-based oxide@polymer composite, which delivered a high capacity of 380 mAh g^{-1} at 1000 mA g^{-1}. X-ray absorption fine structure (XAFS) spectroscopy was employed to probe the electronic environment of Co at the K-edge, while density functional theory (DFT) calculations of the Co d-band density of states (DOS) further revealed the origin of the enhanced charge-storage performance. This integrated strategy provides a powerful pathway for accelerating the discovery of next-generation high-performance electrode materials.

Taner Karagöl*

Istanbul University Medical Faculty, Turkey

Taner Karagöl is a researcher and last-year medical student (Doctor of Medicine) at Istanbul University Medical Faculty. He has collaborated with leading researchers, including Shuguang Zhang at Massachusetts Institute of Technology (MIT). He was a visitor at the Royal Free Hospital London, UCL Medical School for 6 months working with Structural Immunology Group, UCL. With his twin brother Alper Karagöl, their work aims to bridge the disciplines of structural biology, neuroscience, immunology, artificial intelligence, bioinformatics and other fields of biomedical research to advance our understanding of molecular/evolutionary mechanisms and develop innovative therapeutic approaches. Karagöl had worked on water-soluble variants of neurotransmitter transporters and immunomodulatory proteins. By analyzing millions of amino acid substitutions in transmembrane proteins, he maps evolutionary pressures shaping protein function, with major implications for drug resistance and bioengineering. In addition to identifying natural QTY-code substitutions, a key contribution of his work is the identification of truncated isoforms with evolutionary significance. Karagöl has authored multiple peer-reviewed articles on protein evolution, transmembrane structure prediction, and asymmetric mutational dynamics. He has four pending patents on protein design and truncated isoforms. He was admitted to medical school with top-tier distinction and has received several medical educational scholarships and numerous awards, including Dean's Award for Scientific Excellence and Turkish Scientific Research Council International Scientific Meetings Fellowship in 2024. Beyond academic pursuits, he is engaged in scientific advocacy work, focusing primarily on planetary health and public engagement of science.

Computational Data Driven Methods In Discovery Of Native Neurological Transporter Modulators

Taner Karagöl[1], Alper Karagöl[1], Mengke Li[2], Shuguang Zhang[3]

[1]Istanbul University Istanbul Medical Faculty, Istanbul, Turkey

[2]State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic and Developmental Sciences, Shanghai Jiao Tong University, Shanghai, China

[3]Laboratory of Molecular Architecture, Media Lab, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, 02139, USA

The glutamate transporter subfamily (EAAT) is essential for maintaining neurotransmitter homeostasis and its dysregulations have been linked to the etiologies of many conditions, including neurodegenerative diseases. Biological assemblies of glutamate transporters form as cyclic homotrimers, targeting this bio-assembly promises an effective approach for modulatory discovery. Additionally, transcriptional flexibility generates structurally distinct transporter isoforms. In this study, the potential EAAT isoforms were identified through the computational analysis of gene-centric mapping. The conserved features of isoform sequences were revealed by computational chemistry methods alongside AI driven structural predictions from AlphaFold. The isoform complexes were further subjected to a wide range of analyses, 50ns-molecular dynamics simulations, and evolutionary coupling analyses. Accordingly, the root-mean-square-fluctuation, Poisson-Boltzmann Surface-Area, and trajectory analyses were conducted. Our multi-level analysis suggested an inhibitory potential of truncated isoforms on glutamate transporter biological assembly. Moreover, showing the role of transmembrane helices in these interactions, isoforms mimicking the trimerization domain (particularly TM2 helices) demonstrated stronger binding with canonical isoforms. The self-assembly behavior observed in truncated isoforms mimicking canonical TM5 helices indicate a possible protective role against unwanted interactions. Interestingly, EAA2 interactions showed twofold higher binding affinity than EAA1 interactions. Subsequently, focusing on the cross-reactions between EAA1 and isoforms of EAA2, our additional findings also revealed that that EAA2 truncated isoforms form stable complexes with the canonical EAA1. Our computational studies on glutamate transporters not only offer insights into the roles of alternative splicing on neuronal protein interactions, but also identify the critical role of cross-reactivity, and potential drug targets for neurological disorders.

Deepak Chaurasiya*

Indian Institute of Information Technology Allahabad, India

Deepak Chaurasiya received his M.Tech degree in Animal Biochemistry from ICAR–National Dairy Research Institute, Karnal, India, in 2020, PG Diploma in Intellectual Property Rights from Indira Gandhi National Open University, and his B.Tech degree in Biotechnology from Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India, in 2016. He is currently pursuing a Ph.D. degree in Bioinformatics at the Indian Institute of Information Technology Allahabad (an Institute of National Importance), Prayagraj, India. In 2020, he worked as a Junior Project Associate at the Translational Health Science and Technology Institute (THSTI), Faridabad, contributing to SARS-CoV sero-surveillance projects. Since 2023, he has been a Senior Research Fellow in the Biomedical Informatics Lab, Department of Applied Sciences, IIIT Allahabad, where his research focuses on computational biology and machine learning approaches for the study of intrinsically disordered proteins (IDPs). His work includes developing novel predictors for disordered regions, exploring IDP-linked viral proteins, and investigating therapeutic targets in disease-related IDPs. His broader research interests include protein structural biology, molecular modeling, IDP functional annotation, and ML-driven biomedical applications.

Characterising Proteins with AI: Language Models Driving Scientific Discovery

Deepak Chaurasiya[1*]

[1]Biomedical Informatics Lab, Department of Applied Sciences, Indian Institute of Information Technology Allahabad, Prayagraj, Uttar Pradesh-211015, India.

Artificial intelligence is increasingly redefining the way researchers approach molecular characterization, bridging disciplines and automating traditionally experimental tasks. Protein language models (PLMs), trained on massive sequence corpora, exemplify this transformation. Recent advances, such as Fine-tuning protein language models boosts predictions across diverse tasks, show that PLMs can learn contextual biochemical features directly from amino acid sequences, enabling accurate predictions of structure, dynamics, function, and evolutionary conservation. When fine-tuned, PLMs serve as versatile characterisation methods, capable of predicting disordered regions, functional motifs, binding affinities, and mutational impacts at scale. These computational pipelines reduce dependence on costly, time-intensive assays, while providing reproducible and generalizable insights. As a result, PLMs not only automate aspects of protein science but also accelerate discovery in drug development, enzyme engineering, and systems biology. Crucially, the interdisciplinary design of PLMs integrates machine learning, natural language processing, molecular biology, and biophysics, fostering open collaboration and bridging traditionally separate domains of research. Their open-source availability (e.g., ProtBERT, ESM, and TAPE) encourages community adoption and collaborative improvement, making them accessible tools for both computational scientists and experimental researchers. By positioning protein language models as AI-enabled characterization frameworks, this contribution highlights their potential to redefine research automation, democratize access to molecular insights, and drive scientific discovery. Thus, PLMs stand as a frontier where artificial intelligence and biology converge to unlock scalable, interpretable, and impactful pathways in modern science.

Muthuchamy Murugavel*

SRM University Sikkim, Gangtok, INDIA

Dr. Muthuchamy Murugavel is a seasoned chemist and academic leader specializing in porphyrin and corrole chemistry, with a strong focus on the design, synthesis, and photophysical characterization of porphyrinoid systems and their metal complexes. He has built a robust research portfolio spanning the synthesis of organic π-conjugated materials to carbaporphyrins and Multimetallic assemblies, underpinned by deep expertise in coordination chemistry and molecular design. Currently serving as Assistant Professor, Department of Chemistry, School of Basic Sciences, SRM University Sikkim (Gangtok, INDIA).

Machine Learning Approaches to Surpass the Limitations of the Beer-Lambert Law

Many scientific and industrial applications depend on the precise measurement of chemical concentrations. The current study demonstrates how an inventive method of combining photographic images with a machine learning (ML) model successfully estimates the concentration of a chemical compound in solution. A predictive model is created by taking photographs of K2Cr2O7 solutions and evaluating the color intensities of those photographs using a ridge regression model. The prediction precision of the model had been evaluated using 210 images and a high correlation between actual and predicted K2Cr2O7 concentrations was obtained with MAE, MSE, and RMSE of 1.4 × 10-5, 3.4 × 10-10, and 1.0 × 10-5, respectively. The ridge regression model is also extended to predict the concentration of potassium permanganate (KMnO4) and highlights the potential of integrating machine learning techniques with image analysis to accurately quantify the concentration of any chemical species in the solution state. As this model depends solely on the colour intensity of the sample without any molecular interactions, it exceeds the limitations of the Beer-Lambert law. The created machine learning model also minimizes the requirement of substantial expertise and training and hence bridges the gap between experienced and novice analysts.

(1) Ide,Y.; Shirakura, H.; Sano, T.; Murugavel, M.; Inaba, Y.; Hu, S.; Takigawa, I.; Inokuma, Y. Machine Learning-Based Analysis of Molar and Enantiomeric Ratios and Reaction Yields Using Images of Solid Mixtures. Ind. Eng. Chem. Res. 2023, 62(35), 13790-13798.
(2) Sano, T.; Ide, Y.; Tsumori, T.; Ubukata, H.; Takigawa, I.; Kageyama, H.; Inokuma, Y. Hydride Content Control of Perovskite Oxyhydride BaTiO3–xHx Supported by Image-Based Machine Learning. ACS Appl. Eng. Mater. 2024, 2, 10, 2391-2396
(3) Pradhan, S.; Bhattarai, J. S.; Murugavel, M.; Sharma, O. P. Machine Learning Approaches to Surpass the Limitations of the Beer-Lambert Law, ACS Omega, 2025, 10, 16, 16597-16601.

Jiangning Song

Monash University, Australia

Jiangning Song is a Full Professor and Head of the AI-driven Bioinformatics and Biomedicine Laboratory in the Monash Biomedicine Discovery Institute (BDI), Monash University, Australia. He is also an Associate Investigator of the ARC Centre of Excellence in Advanced Molecular Imaging, and an affiliated member of the Monash AI Institute. He has a career total of 351 publications in international journals with ~15100 citations, H-index of 64 and i10-index of 250 in Google Scholar as of 22 September 2025. He has published extensively in top-tier journals, e.g. Cell, Nature Biotechnology, Nature Methods, Lancet Oncology, Nature Machine Intelligence, Nature Computational Science, Nature Sustainability, Nature Communications, Lancet Planetary Health, Science Immunology, Science Advances, Cell Genomics, Cell Reports, Nucleic Acids Research, Genomics Proteomics and Bioinformatics, Briefings in Bioinformatics, and Bioinformatics. He is currently an Associate Editor of several international journals including IEEE Journal of Biomedical and Health Informatics, BMC Bioinformatics, Genomics, Proteomics & Bioinformatics, and BMC Genomic Data. He is motivated to design, develop, and deploy cutting-edge data-driven methodologies and techniques to better understand and address a range of open and challenging scientific questions in biomedicine.

From Machine Learning to Biological Insight: Explainable AI for Decoding Immune Recognition

The integration of machine learning into characterization sciences has revolutionized our ability to extract meaningful patterns from complex, high-dimensional data. However, as we increasingly rely on these powerful models to validate systems and make critical predictions, a fundamental challenge emerges: the "black box" problem of machine learning. This talk will illustrate how we can transform machine learning from a mere analytical tool into a partner for generating interpretable, biologically meaningful insights by focusing on a central challenge in immunology: predicting T cell receptor (TCR) binding to antigens. This process is fundamental to adaptive immunity but has proven immensely difficult to model due to the overwhelming diversity and cross-reactivity of TCRs.

To overcome this, we developed EPACT (Epitope-anchored Contrastive Transfer Learning), a novel deep-learning framework tailored for paired human $CD8^{+}$ TCR data. EPACT's power stems from a unique “divide-and-conquer” strategy that creates co-embeddings of TCRs and their peptide-MHC (pMHC) targets, learning a shared semantic space of immune recognition. Crucially, by fine-tuning EPACT on structural data, we extend its utility beyond mere prediction to mechanistic characterization. The model can quantify interchain distances and identify critical contact residues at the TCR–pMHC interface.

This capability allowed us to decipher the structural basis of TCR cross-reactivity across tumour-associated antigens. Our work exemplifies how explainable AI can automate and deepen biological discovery, offering profound insights for designing safer, more effective TCR-based therapies and accelerating the path toward precision immunotherapies.

Jonathan Stokes

McMaster University, Canada

Jon is an Assistant Professor in the Department of Biochemistry and Biomedical Sciences at McMaster University. His research focuses on the development and application of leading-edge machine learning techniques for novel drug discovery and design tasks, with an emphasis on antibacterial, antiviral, and neuro-oncology agents. Jon is also co-founder and CSO of Stoked Bio.

How should we use AI to help us find new antibiotics

Antibiotic resistance is an unmet global challenge and discovering new antibiotics is hard. In this talk, we will explore the utility and limitations of discriminative and generative machine learning methods for antibiotic discovery and design tasks. We will first discuss well-validated graph-based molecular property prediction algorithms, followed by contemporary molecular fragment-based antibiotic design methods. We will then touch upon the importance of user-friendly methods to discover novel antibacterial agents by biologists and chemists without computational backgrounds. Our goal is to understand what machine learning can and can't do to help make antibiotic discovery maximally efficient and inexpensive.

Jun Jiang

University of Science and Technology of China, China

Prof. Jun Jiang is a distinguished professor of physical chemistry at the University of Science and Technology of China (USTC), within the School of Chemistry and Materials Science. He earned a Ph.D. in Theoretical Chemistry from the Royal Institute of Technology, Sweden, in 2007, and another Ph.D. in Solid State Physics from the Shanghai Institute of Technical Physics, Chinese Academy of Science, in 2008, following a B.S. degree from WuHan University in 2000. He has published more than 150 papers in prestigious journals including Nature Synthesis, Nature Energy, J. Am. Chem. Soc., Angew. Chem. Int. Ed. Dr. Jiang is a recipient of the “National Science Fund for Distinguished Young Scholars in China”, and has won the “Young Theoretical Chemistry Investigator Award of Chinese Chemistry Society”, “Distinguished Lectureship Award of the Chemical Society of Japan 2020”.

Jiang’s research interests focus on the development and application of theoretical chemistry methods and machine learning techniques in chemistry science. By integrating robotic experiments, quantum chemistry simulations, and artificial intelligence guided predictions, his group has developed a data-driven robotic AI-chemist platform, targeting on a wide range of chemistry and material studies such as Photocatalysis, Biochemistry, Photochemistry, Molecular electronics and photonics. His recent works have demonstrated that generative AI combined with spectroscopic descriptors, has the potential to revolutionize chemical material design by overcoming the constraints of conventional atomic-coordinate-based descriptors, leading to a new way to combine generative AI and mobile robots for the intelligent discovery of new chemicals/materials. For more in-depth information about his research works, his personal homepage (http://staff.ustc.edu.cn/~jiangj1/) and dedicated research platforms might provide additional insights.

Building a Global Infrastructure for AI-Driven Innovation

The primary challenge in AI-driven scientific innovation stems from the scarcity and fragmentation of high-quality data. Existing datasets are often limited, skewed toward successful outcomes, and lack comprehensive multi-property characterization. To address this, a federated global infrastructure of cloud-connected, autonomous laboratories has been proposed, serving as decentralized "data power plants" that generate fully characterized, reproducible datasets via robotic experimental workflows.

The technical pathway to realize this vision is structured into five hierarchical levels:

L0 (Process Automation): Focuses on the automated execution of pre-defined experimental protocols.

L1 (Theory-Experiment Iterative Loop): Algorithm-driven closed-loop where experimental feedback refines computational models.

L2 (Large Model-Driven): Employs large language models (LLMs) or multimodal models for high-level cognitive tasks and experimental decision-making.

L3 (Multi-Platform, Multi-Task): This level enables scalability, intelligent collaboration, and cross-domain scheduling across multiple experimental platforms and tasks.

L4 (Autonomous Scientific Discovery): Achieves emergent exploration and autonomous knowledge creation.

Currently, at the L3 level, we have developed a large-model-driven software platform. It coordinates domain-specific small AI models (e.g., quantum chemistry simulations) and robotic experiments. This integration facilitates intelligent scheduling and real-time fine-tuning of pre-trained models based on experimental data, creating a dynamic, synergistic feedback loop. This approach has led to significant breakthroughs in the design and synthesis of novel materials, including high-entropy catalysts, optical polymer films, luminescent COF materials, and multifunctional proteins. Tasks that previously required over a century of exhaustive experimental screening were completed within months.

By building a global infrastructure for AI-driven innovation, this framework transforms isolated scientific efforts into a collaborative, resilient engine for discovery. It democratizes access to high-quality data and breaks down geographical and institutional barriers to innovation, ultimately accelerating industrial-scale scientific and technological advancements.

Felix Strieth-Kalthoff

University of Wuppertal, Germany

Felix is a tenure-track Assistant Professor of Digital Chemistry at the University of Wuppertal. Born and raised in Germany, Felix studied Chemistry at the University of Münster, where he graduated in 2017. After a research stay at the Massachusetts Institute of Technology (2016–2017, with Tim F. Jamison), Felix returned to Münster and obtained his PhD in Chemistry in the group of Frank Glorius. During that time, his research focused on systematic and computer-aided strategies for reaction development in homogeneous (photo-)catalysis. From 2021 to 2024, Felix was a Schmidt Futures "AI in Science" postdoctoral Fellow in the group of Alán Aspuru-Guzik at the University of Toronto, working on self-driving laboratories for chemistry and materials discovery. In 2024, Felix returned to Germany and assumed his current position as Assistant Professor of Digital Chemistry.

Felix is an organic chemist at heart – and his research interests lie in the development of a digital toolbox for chemistry, and its application to the discovery of sustainable catalytic transformations. His research is inherently interdisciplinary, integrating established experimental techniques from organic chemistry with methods from data science, computational chemistry, and artificial intelligence.

Machine Learning Hits the Lab: Practical Experiment Planning for Molecular Discovery

Over the past decade, Machine Learning approaches have found increasing popularity to support the design of experiments for chemical problems. In this talk, I will discuss some of our recent efforts to incorporate Bayesian Machine Learning into experimental workflows. A particular focus will be on addressing data limitations by integrating expert knowledge into ML models. Using case studies from synthetic chemistry and conjugated organic materials discovery, the talk will highlight the opportunities and challenges in ML to support lab-based decisions.

Jie Xu

Argonne National Laboratory, USA

Jie Xu is scientist at Argonne National Lab. Her research focuses on developing a self-driving laboratory that combines artificial intelligence and robotics systems to accelerate the discovery of functional polymers (https://www.anl.gov/cnm/polybot). Jie earned her PhD in chemistry from Nanjing University, specializing in nanoconfined soft matter, and completed postdoctoral training at Stanford in stretchable electronics. She received the Materials Research Society Postdoctoral Award and is named to the MIT Technology Review's global list of Innovators Under 35, Newsweek list of America's Greatest Disruptors as a budding disruptor, and 2023 Polymeric Materials: Science and Engineering Early Investigator Honoree by the American Chemical Society.

Self-driving Laboratory Empowering Innovation in Functional Polymer Discovery

Autonomous laboratory is the integration of high-throughput robotic experimentation and characterization with data-driven model development to guide the search for targeted formulation and processing conditions. While autonomous/self-driving laboratories have made great advances in pharmaceuticals, hard materials and organic small molecules, the development of such systems for polymeric materials is in a relatively nascent stage. We have been developing an autonomous discovery laboratory, named Polybot, that allows us to rapidly go from a polymer material concept to realized manifestations of final, testable materials targeted at relevant properties. In this talk, I will focus on the autonomous discovery of electronic properties of polymers, covering topics from the inverse discovery of electrochromic polymer structures, the controlled assembly of conducting polymers through solution processing, and the discovery of design principles for mixed-conducting polymers in electrochemical transistors. We will also discuss ongoing efforts to evolve Polybot into a more adaptive system with enhanced human-machine interfaces and as a community resource by building a database.

Melodie Christensen

Merck, USA

Melodie Christensen serves as a Director in the Merck Process Research & Development Enabling Technologies Data-Rich Experimentation group. She earned her Master of Science in Chemistry from the American University in 2004 and began her professional journey as a process chemist at Schering-Plough Research Institute in 2005.

In 2010, Melodie joined Merck's Catalysis and Automation group, where she focused on high-throughput experimentation and laboratory automation. Her dedication to continuous learning and professional growth led her to pursue a PhD in Chemistry at the University of British Columbia, which she completed in 2022 under the guidance of Professor Jason Hein. Her doctoral research explored the integration of data science with custom laboratory automation, contributing to advancements in self-driving labs.

Melodie is deeply committed to fostering a collaborative and supportive environment for her team, encouraging them to explore and innovate within the realm of data-rich experimentation. Outside of her professional life, she enjoys yoga, container gardening, flower arranging, and culinary pursuits, which provide her with balance and inspiration.

Advancing Pharmaceutical Discovery Through AI-Driven Experimental Design and Data Architecture Innovation

The proliferation of artificial intelligence and machine learning technologies across the physical sciences has catalyzed a paradigm shift in research methodologies, enabling unprecedented opportunities for research acceleration. Pharmaceutical research organizations are increasingly leveraging these capabilities to transform traditional approaches to drug development and process optimization.

This presentation will examine Merck Process Research and Development's strategic digital transformation initiative, specifically addressing the design, development, and deployment of AI/ML-enabled experimental design platforms within our research ecosystem. We will present our systematic approach to integrating predictive modeling, automated experimental planning, and adaptive optimization algorithms to enhance research efficiency and decision-making accuracy.

Central to this transformation is the establishment of robust, scalable data architectures that ensure seamless access to contextualized, high-quality datasets. Our implementation framework addresses critical challenges in data standardization, provenance tracking, and real-time analytics integration, enabling researchers to harness the full potential of accumulated experimental knowledge.

He Wang*

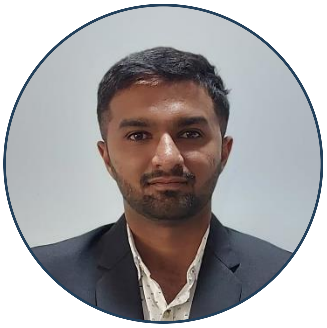
University of Chinese Academy of Sciences (UCAS), China

He Wang received his Ph.D. in Theoretical Physics from Beijing Normal University in 2020. After graduation, he held postdoctoral positions at the Institute of Theoretical Physics, Chinese Academy of Sciences; Peng Cheng National Laboratory (as a visiting scholar); and the University of Chinese Academy of Sciences.
Dr. Wang's research integrates theoretical physics, data science, and artificial intelligence, focusing on gravitational-wave astronomy, machine learning for astrophysical signal processing, and AI-assisted scientific discovery. As a member of the LIGO–Virgo–KAGRA (LVK) collaboration and co-chair of its Machine Learning Working Group (MLA), he contributes to international efforts in data analysis, noise modeling, and algorithmic innovation.
He has published over 30 papers in leading journals and currently plays a key role in the Taiji space-based gravitational-wave mission, leading research on data analysis and machine learning-driven signal processing. Ultimately, his work aims to accelerate scientific discovery by harnessing advanced technologies—including large language models (LLMs)—to deepen our understanding of the universe.

Automated Algorithmic Discovery for Gravitational-Wave Detection Guided by LLM-Informed Evolutionary Monte Carlo Tree Search

He Wang[1,2]* and Liang Zeng[3]*
[1]International Centre for Theoretical Physics Asia-Pacific, University of Chinese Academy of Sciences, 100190, Beijing, China.
[2]Taiji Laboratory for Gravitational Wave Universe, University of Chinese Academy of Sciences, 100049, Beijing, China.
[3]Tsinghua University, 100084, Beijing, China.

Gravitational-wave signal detection with unknown source parameters buried in dynamic detector noise remains a formidable computational challenge. Existing approaches face core limitations from restrictive assumptions: traditional methods rely on predefined theoretical priors, while neural networks introduce hidden biases and lack interpretability. We propose Evolutionary Monte Carlo Tree Search (Evo-MCTS), the first integration of large language model (LLM) guidance with domainaware physical constraints for automated gravitational wave detection.
This framework systematically explores algorithmic solution spaces through tree-structured search enhanced by evolutionary optimization, combining MCTS for strategic exploration with evolutionary algorithms for solution refinement. The LLM component provides domain-aware heuristics while maintaining interpretability through explicit algorithmic pathway generation. Experimental validation demonstrates substantial performance improvements, achieving a 20.2% improvement over stateof- the-art gravitational wave detection algorithms on the MLGWSC-1 benchmark dataset and a remarkable 59.1% improvement over other LLM-based algorithm optimization frameworks. Beyond performance improvements, our framework establishes a transferable methodology for automated algorithmic discovery across computational science domains.



Ujban Hussain*

RTM Nagpur University, India

Ujban Hussain obtained his Master of Pharmacy (M.Pharm) degree in Pharmaceutics from the Department of Pharmaceutical Sciences, Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur, India, in 2025. His research focuses on the integration of artificial intelligence, nanotechnology, and translational pharmaceutics to accelerate drug discovery and improve therapeutic delivery, particularly in central nervous system (CNS) drug administration. Since commencing his research career in 2024, he has co-authored over twenty-five international publications in high-impact Q1–Q2 journals and presented his work at multiple national and international scientific conferences. His contributions have been recognized through more than a dozen academic and innovation awards across twelve conferences, including distinctions from the VIDC, ASTHIBPS 2024, and the Smart India Hackathon. His broader research interests encompass AI-driven research automation, computational pharmaceutics, and data-guided strategies for advanced drug delivery design.

Ann-Powered Research Automation: A Computationally Engineered Platform For Targeted Intranasal Drug Delivery

Ujban Hussain[1], Veena Belgamwar[1], Prashant Puranik[1], Sameer Sheikh[1], Satyendra Prasad[1]

[1]Department of Pharmaceutical Sciences, Rashtrasant Tukadoji Maharj Nagpur University, Nagpur, India

Artificial neural networks (ANN) are increasingly central to research automation, offering predictive power that accelerates discovery while reducing costly trial-and-error experimentation. We present an ANN-driven computational pipeline, integrated with experimental validation, to demonstrate how predictive modeling can automate drug delivery system development. The platform employs ANN-enhanced pharmacokinetic simulations, coupled with X-Booster algorithms, to predict drug transport and absorption pathways. As a proof-of-concept, the system was applied to intranasal CNS delivery, where predictions guided the formulation of chitosan-based nanocarriers optimized for mucoadhesion, stability, and brain targeting. Experimental validation through in-vitro uptake and in-vivo pharmacokinetics confirmed close agreement with ANN forecasts, achieving encapsulation efficiency above 85% and sustained brain distribution within 2% of predicted values. This alignment illustrates how ANN-enabled pipelines streamline formulation optimization, reduce translational risk, and accelerate therapeutic development. By embedding neural networks into discovery and validation workflows, this work highlights the transformative role of ANN in automating research and advancing next-generation therapeutic strategies.

Jona Kayser*

Max Planck Zentrum für Physik und Medizin, Germany

Dr. Jona Kayser is an Emmy Noether Research Group Leader at the Max Planck Zentrum für Physik und Medizin (MPZPM) in Erlangen, Germany. His group bridges physics and medicine using artificial intelligence to understand and steer the evolution and therapy dynamics of solid tumors. By integrating quantitative imaging, cellular biophysics, and machine learning, his research aims to reveal emergent physical mechanisms, such as collective cell motion and spatial organization, shape therapy resistance and to design adaptive treatment strategies guided by physics-based AI models. Before establishing his independent group, Jona was a DFG Research Fellow at the University of California, Berkeley, and completed his PhD in biophysics at the Technical University of Munich.

Reinforcement Failing guides the discovery of emergent physical dynamics in adaptive tumor therapy

Artificial intelligence is revolutionizing scientific discovery in medicine, with reinforcement learning (RL) emerging as a promising tool for optimizing therapeutic strategies. Yet applying RL to complex scenarios such as therapy dynamics in solid tumors is constrained by the challenge of constructing training environments that are both computationally efficient and mechanistically interpretable. Here we introduce Reinforcement Failing, an AI-guided, human-in-the-loop discovery framework that shifts the focus from agent policy optimization to the refinement of the training environment itself. By combining multi-fidelity RL with group-relative performance evaluation across agent cohorts, Reinforcement Failing systematically reveals emergent mechanisms that first-principles models overlook. We apply this framework to adaptive therapy in solid tumors, which seeks to delay resistance-mediated treatment failure. In this setting, Reinforcement Failing uncovered a coupling between the mechanically driven collective motion of cells and spatially-heterogeneous proliferation that strongly influences therapy outcomes. Incorporating these emergent physical mechanisms into an augmented training environment improved cross-environment therapeutic performance and exposed potential pitfalls in translation.
More broadly, these findings position Reinforcement Failing as a powerful artificial scientific discovery framework, capable of deciphering high-complexity processes at the interface of physics, machine learning, and medicine.

SPEAKER LIST

Abstracts appear in this book in presentation order. Below is an alphabetical list of speakers and the corresponding page for their printed abstracts. The scientific organizers have selected the noted Selected Talk (ST) presenters, who were chosen from our abstract submissions.