

## Abstracts of Invited Speakers

Updated on March 10, 2025.  
This file will be updated regularly.

# Fueling the Digital Chemistry Revolution with Language and Multimodal Foundation Models

**Teodoro Laino,<sup>1\*</sup>**

<sup>1</sup>IBM Research

[\\*teo@zurich.ibm.com](mailto:*teo@zurich.ibm.com)

One of the most important outcomes of organic chemistry is the creation of newly designed molecules. The application of domain knowledge gained through decades of laboratory experience has been critical in the synthesis of many new molecular structures. Nonetheless, most synthetic success stories are preceded by lengthy periods of unfruitful explorations. While automation systems proved exceptional in specific fields such as high-throughput chemistry, their use in general-purpose workflows remains a highly complex task, requiring the development of always unique software codifying distinct types of chemical operations. The digital revolution in chemistry hopes to streamline the adoption of digital models and automation with the use of data.

In the last years, natural language processing models have emerged as one of the most effective, scalable approaches for capturing human knowledge and modelling chemical processes in organic chemistry. Its use in machine learning tasks demonstrated high quality and ease of use in problems such as predicting chemical reactions [1-2], retrosynthetic routes [3], digitizing chemical literature [4], predicting detailed experimental procedures [5], designing new fingerprints [6] and yield predictions [7]. In this talk, I'll talk about the use and impact of language and multimodal foundation models in chemistry by highlighting the critical role of NLP architectures in implementing the first cloud-based AI-driven autonomous laboratory [8] and the use of visual transformer for implementing an efficient data capturing scheme.

## References:

- [1] IBM Research Europe, *Chem. Sci.*, 2018, **9**, 6091-6098
- [2] IBM Research Europe, *ACS Cent. Sci.* 2019, **5**, **9**, 1572-1583
- [3] IBM Research Europe, *Chem. Sci.*, 2020, **11**, 3316-3325
- [4] IBM Research Europe, *Nat. Comm.*, 2020, **11**, 3601
- [5] IBM Research Europe, *Nat. Comm.*, 2021, **12**, 2573
- [6] IBM Research Europe, *Nat. Mach. Intel.*, 2021, **3**, 144-152
- [7] IBM Research Europe, *Mach. Learn.: Sci. Technol.*, 2021, **2**, 015016
- [8] <https://rxn.res.ibm.com>
- [9] IBM Research Europe, *Digital Discovery*, 2025



Teodor Laino received the Master degree in theoretical chemistry in 2001 (University of Pisa and Scuola Normale Superiore di Pisa, Italy) and the doctorate in computational chemistry in 2006 (Scuola Normale Superiore di Pisa, Italy) defending a thesis on 'Multi-Grid QM/ MM Approaches in ab initio Molecular Dynamics' supervised by Prof. Dr. Michele Parrinello. From 2006 to 2008, Teo worked as a post-doctoral researcher in the research group of Prof. Dr. Jürg Hutter at the University of Zurich, contributing to the development of the CP2K simulation package. In 2008, Teo joined the IBM Research - Zurich Laboratory (ZRL) as Research Scientist. He is currently Distinguished Research Scientist and manager.

His research interests focus on developing machine learning/artificial intelligence technologies to digitalize chemistry and materials science, with [IBM RXN for chemistry](#) being an example of a recent community success. In 2022, the team received the Sandmeyer Award of the Swiss Chemical Society for the important contributions to the field of digital chemistry.

# Four Generations of Neural Network Potentials

Jörg Behler

Lehrstuhl für Theoretische Chemie II, Ruhr-Universität Bochum, Germany

and

Research Center Chemical Sciences and Sustainability, Research Alliance Ruhr, Germany

In recent years, there has been tremendous progress in the development of machine learning potentials (MLP) employing many different approaches [1]. Neural network potentials (NNPs), which have been introduced more than two decades ago, are an important class of MLPs. In general, MLPs can be classified into different generations representing their evolution with respect to applicability and physical phenomena they are able to describe. While the first generation of MLPs has been restricted to small molecules with only a few degrees of freedom, the second generation extended their applicability to high-dimensional systems containing thousands of atoms by constructing the total energy as a sum of environment-dependent atomic energies [2]. Long-range electrostatic interactions can be included in third-generation MLPs employing environment-dependent charges [3], and in the following years this locality approximation could be overcome by the introduction of fourth-generation MLPs [4,5], which are able to describe non-local charge transfer using a global charge equilibration step. In this talk an overview about the rapid evolution of high-dimensional NNPs will be given along with typical applications in chemistry and materials science.

[1] J. Behler, J. Chem. Phys. **145** (2016) 170901.

[2] J. Behler and M. Parrinello, Phys. Rev. Lett. **98** (2007) 146401.

[3] N. Artrith, T. Morawietz, J. Behler, Phys. Rev. B **83** (2011) 153101.

[4] S. A. Ghasemi, A. Hofstetter, S. Saha and S. Goedecker, Phys. Rev. B **92** (2015) 045131.

[5] T. W. Ko, J. A. Finkler, S. Goedecker, J. Behler, Nature Comm. **12** (2021) 398.

**Keywords:** Machine Learning Potentials, Molecular Dynamics, Potential Energy Surface

---

## Biography :



Jörg Behler graduated in chemistry at the University of Dortmund in 2000. In 2004 he obtained his PhD at the Fritz-Haber-Institute in Berlin. After a postdoctoral stay at the ETH Zürich, in 2007 he established his own research group at the Ruhr-Universität Bochum funded by a Liebig, an Emmy Noether and a Heisenberg fellowship. In 2013 he received the Hans G. A. Hellmann award for his work on the development of high-dimensional neural network potentials. In 2017 he became a full professor for theoretical chemistry at the University of Göttingen. In 2022 he returned to Bochum for a research professorship at the newly founded Research Center Chemical Sciences and Sustainability and for establishing a new Chair for Theoretical Chemistry II. His main research interest is the development and application of machine learning potentials in chemistry and materials sciences.

# Deep Scopie: an introduction to machine learning for imaging

David Rousseau ; [david.rousseau@univ-angers.fr](mailto:david.rousseau@univ-angers.fr)

Université d'Angers, LARIS, UMR INRAe IRHS, Angers, France.

We introduce machine learning (both shallow and deep learning) in the context of imaging. We specifically address challenges related to computational imaging where the raw acquisition requires processing before being interpretable by humans or machines. We illustrate this topic with various applications such as denoising, super-resolution, inverse problems, multimodal imaging, smart scanning... We will discuss the opportunities and challenges of bridging the model-based physics of imaging with a purely data-driven approach. The talk will be illustrated with a wide range of imaging modalities such as computed tomography imaging spectrometer [1], fluorescent microscopy [2], Raman microscopy [3], X-Ray tomography [4], RGB-Depth imaging [5].

David Rousseau serves as professor of data science in the physics department at the Université d'Angers, where he leads the Imaging for Horticulture and Phenotyping team.

## A selection of recent publications on the topic by the speaker:

- [1] Douarre, C., Crispim-Junior, C. F., Gelibert, A., Germain, G., Tougne, L., & Rousseau, D. (2021). CTIS-Net: a neural network architecture for compressed learning based on computed tomography imaging spectrometers. *IEEE Transactions on Computational Imaging*, 7, 572-583.
- [2] Ahmad, A., Sala, F., Paiè, P., Candeo, A., D'Annunzio, S., Zippo, A., ... & Rousseau, D. (2022). On the robustness of machine learning algorithms toward microfluidic distortions for cell classification via on-chip fluorescence microscopy. *Lab on a Chip*, 22(18), 3453-3463.
- [3] Gilet, V., Mabillean, G., Loumagne, M., Coic, L., Vitale, R., Oberlin, T., ... & Rousseau, D. (2023). Superpixels meet essential spectra for fast Raman hyperspectral microimaging. *Optics Express*, 32(1), 932-948.
- [4] Hamdy, S., Charrier, A., Corre, L. L., Rasti, P., & Rousseau, D. (2024). Toward robust and high-throughput detection of seed defects in X-ray images via deep learning. *Plant Methods*, 20(1), 63.
- [5] F. Mercier, G. Couasnet, A. El Ghaziri, N Bouhlef, A Sarniguet, M Marchi, M Barret, D Rousseau. Deep-learning-ready RGB-Depth images of seedling development. *Plant Methods* (in press 2025)

**Youtube channel** : <https://www.youtube.com/@imhorphenbioimagingresearch95>

**Keywords:** Computational imaging ; Compressed sensing, Deep learning.



# Reinforcement learning for bio-inspired navigation in complex environments

Aurore Loisy

Aix Marseille Univ, CNRS, Centrale Med, IRPHE, Marseille, France

Navigation is about optimising a route to go from A to B. Animal and robotic navigation however fundamentally differ from the routing of planes and ships, because it is autonomous : the self-propelled « agent » has only access to information from its own sensors to make decisions. This type of problem is well-suited for reinforcement learning, a branch of artificial intelligence which has gained popularity by beating human players at games. I will show how we can leverage modern (deep) reinforcement learning techniques to solve navigation problems inspired by the animal world, such as the vertical migration of plankton through the water column, the search for an odor source by insects, and chemical-guided collective foraging.

**Keywords:** reinforcement learning, partially observable Markov decision process, olfactory search, navigation in turbulent flows, bio-inspiration

# Encoding and Decoding Chemistry with Language Models

Kevin Maik Jablonka

Friedrich Schiller Universität Jena & Helmholtz Institute for Polymers in Energy Applications Jena

The field of chemical sciences has seen significant advancements with the use of data-driven techniques, particularly with large datasets structured in tabular form.

However, collecting data in this format is often challenging in practical chemistry, and text-based records are more commonly used [1]. Using text data in traditional machine-learning approaches is also difficult.

Recent developments in applying large language models (LLMs) to chemistry have shown promise in overcoming this challenge. LLMs can convert unstructured text data into structured form and can even directly solve predictive tasks in chemistry. [2, 3] In my talk, I will present the impressive results of using LLMs, showcasing how they can autonomously utilize tools and leverage structured data and “fuzzy” inductive biases.

To enable the training of a chemical-specific large language model, we have curated a new dataset along with a comprehensive toolset to utilize datasets from knowledge graphs, preprints, and unlabeled molecules. To evaluate frontier models trained on such a dataset, we specifically designed a benchmark to evaluate the chemical knowledge and reasoning abilities. I will present the latest results, demonstrating the potential of LLMs in advancing chemical research. [4]

## References:

- [1] Jablonka, K. M.; Patiny, L.; Smit, B. *Nat. Chem.* 2022, 14 (4), 365–376.
- [2] Jablonka, K. M.; et al. *Digital Discovery* 2023, 2 (5), 1233–1250.
- [3] Jablonka, K. M.; Schwaller, P.; Ortega-Guerrero, A.; Smit, B. *Leveraging large language models for predictive chemistry. Nat. Mach. Int.* 2024, 6, 161–169.
- [4] Mirza, A.; Alampara, N.; Kunchapu, S.; Emoekabu, B.; Krishnan, A.; Wilhelmi, M.; Okereke, M.; Eberhardt, J.; Elahi, A. M.; Greiner, M.; Holick, C. T.; Gupta, T.; Asgari, M.; Glaubitz, C.; Klepsch, L. C.; Köster, Y.; Meyer, J.; Miret, S.; Hoffmann, T.; Kreth, F. A.; Ringleb, M.; Roesner, N.; Schubert, U. S.; Stafast, L. M.; Wonanke, D.; Pieler, M.; Schwaller, P.; Jablonka, K. M. *Are Large Language Models Superhuman Chemists?* arXiv 2024. <https://doi.org/10.48550/ARXIV.2404.01475>.

**Keywords:** Machine Learning, LLM, Benchmark



Kevin Jablonka leads an independent research group at the Helmholtz Institute for Polymers in Energy Applications of the University of Jena and the Helmholtz Center Berlin where we focusses on designing materials that work in the real world using data-driven techniques. He belongs to a new generation of scientists with a broad skill set, combining expertise in chemistry, materials science, and artificial intelligence. Recently, Kevin has been at the forefront of applying Large Language Models to chemistry and materials science.