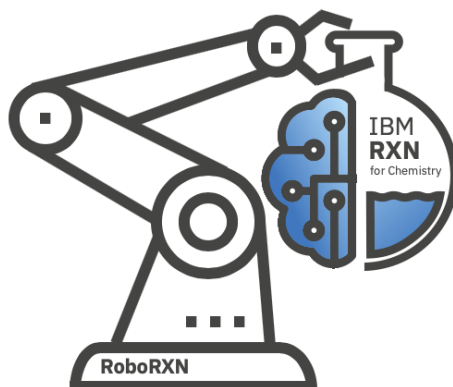


Fueling the Digital Chemistry Revolution with Language Models

Teodoro Laino, Alain Vaucher, Matteo Manica,
Alessandro Castrogiovanni, Antonio Cardinale, Joppe Geluykens, Aleksandros Sobczyk,
Philippe Schwaller, Alessandra Toniato, Heiko Wolf, Theophile Gaudin, Federico Zipoli

IBM Research Europe, Säumerstrasse 4, CH-8803 Rüschlikon (Switzerland)
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 impact of language models in chemistry by highlighting the critical role of NLP architectures in implementing the first cloud-based AI-driven autonomous laboratory [8].

- [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>