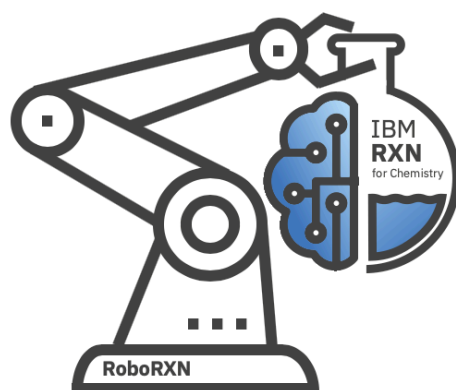


The Power of Language in Chemistry (and more)

Teodoro Laino, Alain Vaucher, Matteo Manica,
Alessandra Toniato, Philippe Schwaller, Yves Gaetan Nana Teukam,
Loic Kwate Dassi, Andrea Byekwaso, Alessandro Castrogiovanni

IBM Research Europe, Säumerstrasse 4, CH-8803 Rüschlikon (Switzerland)
teo@zurich.ibm.com

Natural language processing models in chemistry have emerged as one of the most effective approaches for capturing human knowledge and modelling creativity in organic chemistry. Its application 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 presentation, I will discuss the impact of language models in chemistry by describing first the implementation of the first cloud-based AI-driven autonomous laboratory [8].



Finally, I will focus on the most recent applications of language models to address the problem of characterizing unknown enzymes, developing of human-in-the-loop schemes for retrosynthetic strategies or promoting sustainability and green chemistry strategies with ad-hoc AI models.

- [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>