nature machine intelligence

Article

https://doi.org/10.1038/s42256-024-00832-8

Augmenting large language models with chemistry tools

Received: 13 September 2023

Accepted: 27 March 2024

Published online: 8 May 2024

Check for updates

Andres M. Bran[1,2,6], Sam Cox[3,4,6], Oliver Schilter[1,2,5], Carlo Baldassari[5], Andrew D. White[3,4] & Philippe Schwaller[1,2]

Large language models (LLMs) have shown strong performance in tasks across domains but struggle with chemistry-related problems. These models also lack access to external knowledge sources, limiting their usefulness in scientific applications. We introduce ChemCrow, an LLM chemistry agent designed to accomplish tasks across organic synthesis, drug discovery and materials design. By integrating 18 expert-designed tools and using GPT-4 as the LLM, ChemCrow augments the LLM performance in chemistry, and new capabilities emerge. Our agent autonomously planned and executed the syntheses of an insect repellent and three organocatalysts and guided the discovery of a novel chromophore. Our evaluation, including both LLM and expert assessments, demonstrates ChemCrow's effectiveness in automating a diverse set of chemical tasks. Our work not only aids expert chemists and lowers barriers for non-experts but also fosters scientific advancement by bridging the gap between experimental and computational chemistry.

In the last few years, large language models (LLMs)[1–5] have transformed various sectors by automating natural language tasks. A prime example of this is the introduction of GitHub Copilot in 2021[6] and more recently StarCoder[7], which provides proposed code completions based on the context of a file and open windows and increases developers' productivity[8]. Most recent advances are based on the Transformer architecture[9], introduced for neural machine translation and extended to various natural language processing tasks demonstrating remarkable few-shot and zero-shot performance[2]. Nevertheless, it is crucial to recognize the limitations of LLMs, which often struggle with seemingly simple tasks like basic mathematics and chemistry operations[10,11]. For instance, GPT-4 (ref. 12) and GPT-3.5 (ref. 13) cannot consistently and accurately multiply 12,345 × 98,765 or convert IUPAC names into the corresponding molecular graph[14]. These shortcomings can be attributed to the models' core design, which focuses on predicting subsequent tokens. To address these limitations, one viable approach is to augment LLMs with dedicated external tools or plugins, such as a calculator for mathematical operations or OPSIN[15] for IUPAC-to-structure conversion. These specialized tools provide exact answers, thereby compensating for the inherent deficiencies of LLMs in specific domains and enhancing their overall performance and applicability.

Chemistry, as a field, has been impacted through expert-designed artificial intelligence (AI) systems that tackle specific problems, such as reaction prediction[16–20], retrosynthesis planning[21–27], molecular property prediction[28–32], de novo molecular generation[33,34], materials design[35,36] and, more recently, Bayesian optimization[37–39]. Due to the nature of their training data, it has been shown that code-generating LLMs do possess some understanding of chemistry[14], allowing them to adapt to observations, plan over multiple steps and respond correctly to intent in a chemical setting[13,40–44]. Still, the automation levels achieved in chemistry remain relatively low compared to other domains, primarily due to its highly experimental nature, the lack of data and the limited scope and applicability of computational tools, even within their designated areas[45].

Integrating such tools tends to occur within isolated environments, such as RXN for Chemistry[18,24,46–48] and AIZynthFinder[25,49,50],

[1]Laboratory of Artificial Chemical Intelligence (LIAC), ISIC, EPFL, Lausanne, Switzerland. [2]National Centre of Competence in Research (NCCR) Catalysis, EPFL, Lausanne, Switzerland. [3]Department of Chemical Engineering, University of Rochester, Rochester, NY, USA. [4]FutureHouse, San Francisco, CA, USA. [5]Accelerated Discovery, IBM Research – Europe, Rüschlikon, Switzerland. [6]These authors contributed equally: Andres M. Bran, Sam Cox. e-mail: andrew@futurehouse.org; philippe.schwaller@epfl.ch