
开源科学大模型“浦科化学”助力科学研究

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/25989.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

开源科学大模型“浦科化学”助力科学研究。1月26日，上海人工智能实验室（上海AI实验室）开源发布首个科学大模型浦科化学（ChemLLM），拓展了大模型助力科学研究的探索路径。

基于书生·浦语2.0强大的基座模型能力，浦科化学通过注入海量化学专业数据，使大模型获得了理解和处理化学相关专业任务的知识与能力。同时，研究人员发现，随着化学能力的获得，大模型的数理学科及推理能力也得到增强。

浦科化学现已开源，并提供免费商用。

开源链接：<https://huggingface.co/AI4Chem/ChemLLM-7B-Chat>

专注核心：化学知识注入，专业能力突出

为了评估浦科化学的专业能力，上海AI实验室的研究人员对其在三种相关任务上的表现进行了测试：分子名称转换、分子性质预测和反应产物预测——这些任务分别涉及化学物质的表示、性质和转化，是化学研究的基础和核心。

分子名称转换方面，要求模型能够在不同分子表示方式之间进行准确转换，如SMILES、IUPAC名称、分子式等；

分子性质预测方面，要求模型能根据分子的结构和组成，预测其化学性质，如沸点、密度、溶解度等；

反应产物预测方面，要求模型能根据给定的反应物和反应条件预测反应产物的结构。

测试结果显示，浦科化学在这三项任务上均表现出色，分别获得22.0、49.0和7.0的分数，远超过其他同规模模型的水平，并超越了GPT-3.5，证明浦科化学在理解和处理化学知识方面具备强大能力。

依托书生·浦语2.0基座模型优秀的多语言能力，浦科化学在经过专业化学知识训练后，还具备了优秀的化学专业中英文翻译能力，可帮助化学研究者跨越语言障碍，准确地翻译化学文献中的专有名词，获取更多的化学知识。

下图展示了浦科化学翻译的一篇2024年1月16日发表在《自然·化学》杂志上论文的摘要。

在专业化学知识训练之外，浦科化学也进行了初高中知识的学习。在回答初高中化学题目时，不仅能给出答案，还能给出具体的解释，下图展示了例子：

齐头并进：逻辑推理增强，泛理科能力涌现

专项能力的增强，是否以牺牲其他能力为代价？实践证明，浦科化学不仅在化学领域表现出色，其在数学、物理及推理等方面能力均得到增强。

研究人员使用MMLU（大规模多任务语言理解数据集）作为测试集来评估浦科化学在通用场景下的性能，对语言模型的多项任务能力进行测试评估，并将浦科化学与其他同等规模的模型进行比较，以评估化学训练数据对基础模型的影响。

评测结果显示，浦科化学在大学化学任务上得分为47.0，超越了所有其他模型，这体现了它在回答化学问题及处理复杂化学任务的能力。

研究人员发现，尽管没有对浦科化学进行数学和物理学科的训练，模型在大学数学、大学物理、STEM（科学、技术、工程和数学）类等相关领域上也获得了最佳成绩。同时，浦科化学在形式

逻辑任务上也达到了最高水平。上述结果表明，当对大模型进行进行化学专业训练时，其数理及推理能力也能得到增强。

有趣的是，浦科化学在道德场景、人文科学类、社会科学类等分项任务上也展现了优异的平均性能，这说明，对大模型进行专业某个学科的知识训练，不仅不会让大模型分心，反而有助于其道德水平及一般任务水平的提升。

举一反三：扩展大模型应用，助推科研新范式

对大语言模型进行化学专项训练，不仅扩展了大模型的应用空间，更为AI for Science相关研究开启了新的探索路径。

上海AI for Science团队面向化学、物理、生命、地球等科学领域，通过深入研究各学科基础理论，结合最新人工智能理论，探索AI驱动重大科学问题的研究范式，加速人工智能在化学、药物研发、新材料、气象等领域的渗透与落地，赋能各行业发展。其中，AI for Chemistry方面的研究以语言模型为核心，通过大模型连接智能化实验设备，全方位提升实验效率，从而实现化学研究的自动化和智能化。相关研究范式的创新，将助推科学发现速度，实现更大的社会效益。

研究团队在化学科研方面取得了一系列科研成果，并于近期发表在《自然-催化》、《美国化学会志》、《德国应用化学》、《中国科学：化学》等期刊上，依托于浦科化学模型，未来会有一系列成果发布，人工智能可为化学研究提供智能化辅助，如化学合成路径规划、化学反应条件优化、实验结果自动化分析等，从而提升化学研究的效率和质量。

相关论文信息：<https://doi.org/10.1038/s41929-023-01032-0>

<https://doi.org/10.1021/jacs.3c01040>

<https://doi.org/10.1002/anie.202308041>

<https://doi.org/10.1007/s11426-023-1812-x>

<https://doi.org/10.1016/j.fmre.2023.03.016>

作者：李玉强等 来源：《自然-催化》

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发