

人工智能驱动的材料化学研究 国际发展态势分析[☆]

吕凤先^{1*} 刘小平^{1,2} 陶冶宇^{1,2}

1 中国科学院文献情报中心 北京 100190

2 中国科学院大学经济与管理学院信息资源管理系 北京 100190

*通信作者 E-mail: lyufx@mail.las.ac.cn

[摘要] [目的/意义] 人工智能驱动的科学范式发生变革，材料化学是其中的一个重要领域。[方法/过程] 针对人工智能驱动的材料化学领域进行研究，调研主要国家在该领域的项目部署，结合WoS数据库中该领域的论文，利用DDA和ITGinsight等分析工具综合分析人工智能驱动的材料化学研究领域的主要国家、重要发文机构、重要研究进展和面临的挑战。[结果/结论] 人工智能驱动的材料化学研究受到美国、欧洲、英国和日本等主要国家/地区的重视，是近年来快速发展的新兴领域。人工智能技术不仅加速了材料的结构化学研究，还推动了材料制备过程的热力学研究和动力学研究，人工智能技术正在为材料化学研究带来前所未有的创新活力，也在加快推进该领域深入发展，有助于解决该领域的复杂问题，尤其是在加快先进材料的发现方面具有重要作用。未来，人工智能驱动的材料化学研究仍有赖于人工智能与材料化学的深度融合。

[关键词] 人工智能 材料化学 重要研究进展 挑战

DOI: 10.15978/j.cnki.1673-5668.20240918

0 引言

材料化学的主要研究内容^[1]包括3个方面：1) 材料的结构化学，即以化学为基础来设计具有特定物理化学特性（如磁性、光学或催化特性）的材料；2) 材料制备过程的热力学，主要研究气体和溶液的热力学、吉布斯自由能变化、相图和相变；3) 材料制备过程的动力学，主要研究传质理论、化学反应速率和化学反应的

控制步骤。

人工智能（AI）正在推动科学范式发生变革，材料化学是其中的一个重要领域。美国、日本、欧洲等主要国家/地区支持材料基因组、智能材料、信息集成型材料等计划，这些计划中包含了对AI驱动的材料化学研究的项目部署。

本文基于AI驱动的材料化学研究的项目和论文数据，结合定性和定量的方法研究如下问题：1) 该领域

[☆]资助项目：国家自然科学基金项目(22342011)。



取得哪些重要进展? 2) 哪些具体的 AI 技术在材料化学中最有影响力? 3) AI 技术在哪些方向加速材料化学领域的研究进程? 4) 该领域面临什么挑战?

1 美国、欧盟、英国、日本等部署项目支持 AI 驱动的材料化学研究

美国通过材料基因组计划在 AI 驱动的材料化学研究领域部署项目。2011 年, 美国推出材料基因组计划, 支

持发现、制造和部署先进材料, 材料基因组计划主要资助机构包括美国国防部、能源部、国家标准与技术研究所 (NIST)、国立卫生研究院 (NIH) 和国家科学基金会 (NSF)。在人工智能驱动材料化学领域, NSF 的主要计划包含“设计材料以革命和设计我们的未来” (DMREF) 和材料创新平台 (MIP) (表 1), NIH 重点支持生物材料相关研发, NIST 重点支持材料数据、工具和技术研发, 以及多个长度和时间尺度的材料现象模拟。

表 1 NSF 的 DMREF 计划和 MIP 计划资助的部分项目

项目名称	机构名称	资助金额 / 万美元	AI 驱动的材料化学研究内容
生物聚合物、自动化细胞基础设施、流动和集成化学: 材料创新平台	加州大学圣塔巴巴拉分校	2022	1) 计算机建模和机器学习相结合 2) 从酵母、真菌和细菌中生产具有立体特异性官能团的单体, 这些单体被送入基于材料基因组计划的分层计算、自动聚合和流动化学循环。模拟工具用于预测合成材料的性能, 并系统地探索新的手性、区域选择性和功能性生物源单体的设计图景 3) 与加州大学洛杉矶分校合作
自动化合理设计的糖材料的合成	弗吉尼亚理工大学	1906	1) 开放式数据库, 分子建模 2) 为用户提供多个开放式数据库和在线服务, 以促进质谱和核磁共振光谱分配的自动化, 提供分子建模的访问, 并实现自动化聚糖合成的优化。 3) 与佐治亚大学、布兰迪斯大学、伦斯勒理工学院和北卡罗来纳大学等机构开展合作
二维晶体联盟	宾夕法尼亚州立大学帕克分校	1341	1) 知识图谱 2) 新的合成能力包括以极高化学计量精度为目标的双坩埚布里奇曼方法, 以及用于可扩展合成空气稳定二维极性金属的集成等离子体/蒸发工具。该合成网络与一套全面的先进光谱、扫描探针和处理工具相互连接, 允许在保持真空或惰性气体环境的同时进行样品交换。整个设施生成的数据被输入到一个示例数据跟踪系统中, 该系统嵌入了知识图谱功能, 以实现数据驱动的材料发现的新模式。
原子尺度数据与涌现的量子积分	康奈尔大学	240	1) 用于解决数据问题的机器学习工具套件 2) 在量子化学推理的指导下, 无机晶体结构数据库和材料项目的数据将用于发现新的描述符, 并预测新的拓扑材料。
基于数据驱动实验集成和多尺度建模, 加速发展铝合金	约翰霍普金斯大学	约 206	利用现代数据科学工具开发数据框架和基础设施, 在复杂的材料开发问题中无缝集成实验和模型, 促进先进材料快速应用于工程应用。
用于光子集成电路的磁-电-光耦合混合超材料薄膜平台	普渡大学	约 200	识别具有超低能量势垒的新型软磁赫斯勒合金, 在磁性隧道结中设计基于随机低势垒磁体的概率比特 (p 比特), 实现概率计算的突破。研究方法: 理论建模、高通量材料筛选、计算模拟、实验合成和表征以及模块化模拟
基于数据科学驱动的多分辨率实验和模拟, 发现高温、抗氧化、复合、浓缩的合金		约 174	旨在优化难熔复合浓缩合金的成分, 以实现高温下强度和抗氧化性的无与伦比的组合。研究方法开发一种迭代方法, 在机器学习加速材料发现 (ML-AMD) 框架内结合多保真度和多成本实验以及基于物理的建模



欧盟和英国支持用 AI/ 机器学习驱动的材料化学研究。2014—2024 年，欧盟地平线 2020 计划和欧洲研究委员会资助“百亿亿级材料设计 (MaX)”项目 (850 万欧元)^[2]、“欧洲 AI 按需平台和生态系统 (AI4EU)”项目 (2067 万欧元)^[3]、“材料基因组”项目 (248.67 万欧元)^[4]、“氢键网络结构复杂性的解码、映射和设计：

从水到蛋白质再到聚合物”项目 (150 万欧元)，以及“新材料发现 (NOMAD)”项目及其扩展项目“新材料发现卓越中心 (NOMAD CoE)” (约 1000 万欧元)^[5-6]。2015—2024 年，英国研究与创新署资助“基于计算与实验集成加速材料发现”项目^[7] (665.06 万英镑)、资助建立化学人工智能 (AIchemy) 中心^[8] (表 2)。

表 2 欧盟 AI 驱动材料化学研究部分项目

项目名称	牵头机构	研究内容	年份
百亿亿级的材料设计 (MaX)	意大利国家研究委员会	通过结合量子力学仿真的预测能力以及新兴高性能计算技术的性能、数据容量和能源可持续性，来促进对材料属性的理解和设计，加速新材料的发现。该项目在高性能计算、数据管理和材料模拟方面使用了与 AI 相关的技术和方法，为在处理复杂系统和大量数据时利用 AI 推动材料科学奠定基础	2018—2022
欧洲 AI 按需平台和生态系统 (AI4EU)	法国泰雷兹集团	AI4EU 项目的活动包括：创建和支持一个跨越 28 个国家的大型欧洲生态系统，以促进所有欧洲 AI 参与者 (科学家、企业家、中小企业、行业、资助组织、公民) 之间的合作；设计一个欧洲 AI 按需平台，共享欧洲项目中产生的 AI 资源；通过 AI4EU 平台实施行业主导的试点，展示该平台实现实际应用和促进创新的能力。该项目为推动材料科学发展提供 AI 资源，为 AI 推动材料科学奠定基础	2019—2021
材料基因组	瑞士洛桑联邦理工学院	研究在合成前可靠地预测新型材料的性能	2015—2021
氢键网络结构复杂性的解码、映射和设计：从水到蛋白质再到聚合物		开发机器学习技术并将其应用于原子模拟，并确定控制氢键化合物结构和性能的设计原则	2016—2021
新材料发现，新材料发现卓越中心 (NOMAD CoE)	德国马克斯·普朗克学会弗里茨哈伯研究所和德国洪堡大学	整合欧洲各地材料相关数据，并提供强大的新工具来搜索、检索和管理这些数据，开辟新的高性能计算机会；将计算材料科学提升到超级计算的新水平 (百亿亿级计算)，包括开发百亿亿级算力的算法，以创建真实世界、工业相关、复杂材料的准确预测模型，在水分解和新型热电材料两个用例中测试和演示结果	2015—2018 2020—2024

日本支持 AI/ 机器学习和数据驱动的物质和材料研发。2015 年至今，日本科学技术厅资助日本国立材料科学研究所“信息集成倡议材料研究”项目，开展数据驱动 (信息集成) 的材料研究^[9]。2017 年起，丰田研究所与大学和国家实验室合作，投入 7000 余万美元，支持材料发现基础研究，包括 AI/ 机器学习理解电池退化和指导催化剂纳米颗粒合成^[10-11]。2020 年，日本成立产学研合作组织“材料集成联盟”，推动综合型材料开发系统实际应用，利用 AI 和数据科学，该系统可以预测材料加工工艺、结构和

性能，也可以根据性能需求快速探索最佳加工工艺和化学成分^[12]。2024 年，日本文部科学省提出支持光学与信息、材料等相互融合的尖端技术等战略目标^[13] (表 3)。

2 论文数据来源和分析工具

论文数据来自科睿唯安 Web of Science (WoS) 数据库的 SCI-E 和 CPCI-S 引文数据库。论文检索式是 2 个部分的并集。1) AI 技术。本文按照算法分类关系进



表3 英国 AI 驱动材料化学研究部分项目

项目名称	牵头机构	研究内容	年份
化学人工智能 (AIChem) 中心	利物浦大学和伦敦帝国理工学院	研究基础 AI 方法、实验和计算化学，以及用于化学发现的自主闭环机器人技术	2024 (启动)
基于计算与实验集成加速材料发现	利物浦大学	开发一种设计、合成和评估新功能材料 (有机固体、无机固体，以及单个同时包含有机和无机模块的固体混合物) 的综合方法	2015—2021

行关键词分解,还直接纳入了 WoS“计算科学,人工智能”类别中的全部论文。首先, AI 的主流算法为机器学习,机器学习可分为监督学习、无监督学习和半监督学习等类型(监督学习又可分为回归与分类),其常用算法具体包括神经网络、支持向量机、决策树、随机森林和强化学习等;另一主流算法为深度学习,又可分为卷积神经网络、循环神经网络等;另外,遗传算法也是人工智能应用于材料研发的热点算法。2) 材料化学。本文根据材料化学的主要研究内容制定材料化学的检索式,同时将论文类别限定在既属于 WoS“化学”又属于“材料”类别,或者将论文发表的期刊限定在《自然》《科学》《自然·化学》《自然评论化学》《自然评论材料》,并对研究内容是否属于 AI 驱动的材料化学研究进行人工判读。论文检索式见文后附录,截至 2024 年 9 月,检索到 1991—2024 年相关文章 2338 篇。

本文数据分析工具主要采用德温特数据分析软件 (DDA) 和通用科技文本可视化挖掘系统 (ITGinsight)。

3 结果与分析

3.1 时间分布特征

本文依据年发文量分析该领域论文时间分布特征,以了解 AI 在材料化学领域应用发展的总体趋势。如图 1 所示, AI 驱动的材料化学研究最早的一篇文献始于 1991 年,早期的代表性研究包括将神经网络应用于聚合物材料设计^[14]和有机物吸附^[15]等。2006 年前,该领域论文产出少。自 2007 年开始, AI 驱动的材料化学研究的论文数量整体呈现稳中有升趋势。2016 年至今,该领域文献呈现快速增长态势(图 1),这一转折点恰与“人工智能元年”相互吻合。

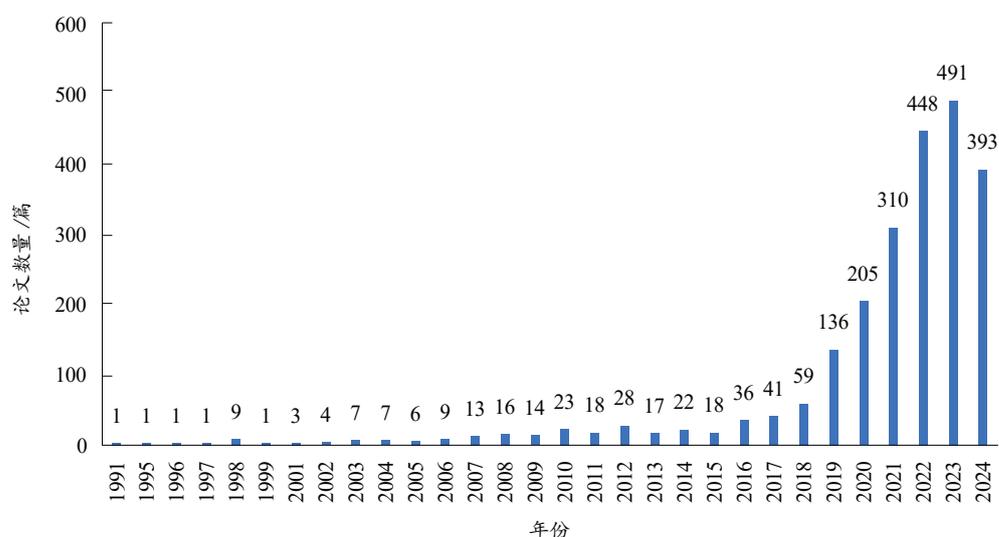


图1 AI 驱动的材料化学研究各年度论文数量



3.2 国别分布特征

本文进行国别发文量统计分析，以了解开展 AI 驱动的材料化学研究的主要国家。如图 2 所示，中国在 AI 驱动的材料化学研究中发文量（863 篇）排名第

1 位；美国发文量（584 篇）排名第 2 位；印度发文量（164 篇）排名第 3 位，图 2 显示了 AI 驱动的材料化学研究领域论文数量排名前 10 位的国家以及具体发文数量。

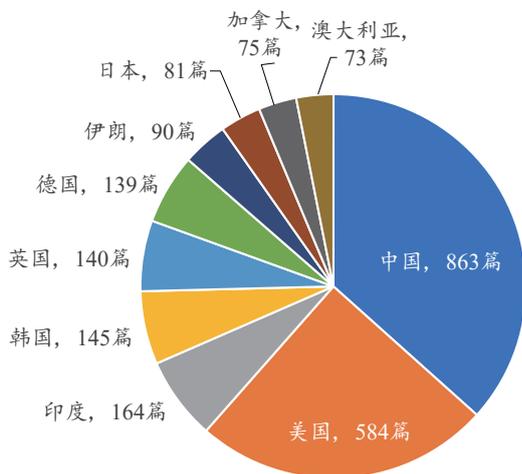


图 2 AI 驱动的材料化学研究领域论文数量 Top10 国家

3.3 研究机构分布特征

对研究机构发文量进行统计分析，以了解 AI 驱动的材料化学研究的关键研究机构。如图 3 所示，在 AI 驱动的材料化学研究领域，论文数量排名前 10 名的机构包括中国科学院、橡树岭国家实验室、西安交通大学、哈尔滨工业大学、西北工业大学、清华大学、麻省理工学院、剑桥大学、佐治亚理工学院和北京科技大学。其中，中国科学院在该领域的发文量（108 篇）排名第 1 位；

橡树岭国家实验室发文量（37 篇）排名第 2 位；西安交通大学发文量（36 篇）排名第 3 位。

同时，本文使用 DDA 对发文数量排名 Top10 机构进行研究机构合作网络分析。如表 4 所示，中国科学院与其他机构合作发文数量最多，合作机构包括美国橡树岭国家实验室、西安交通大学、哈尔滨工业大学、西北工业大学、清华大学、麻省理工学院、剑桥大学、佐治亚理工学院和北京科技大学。

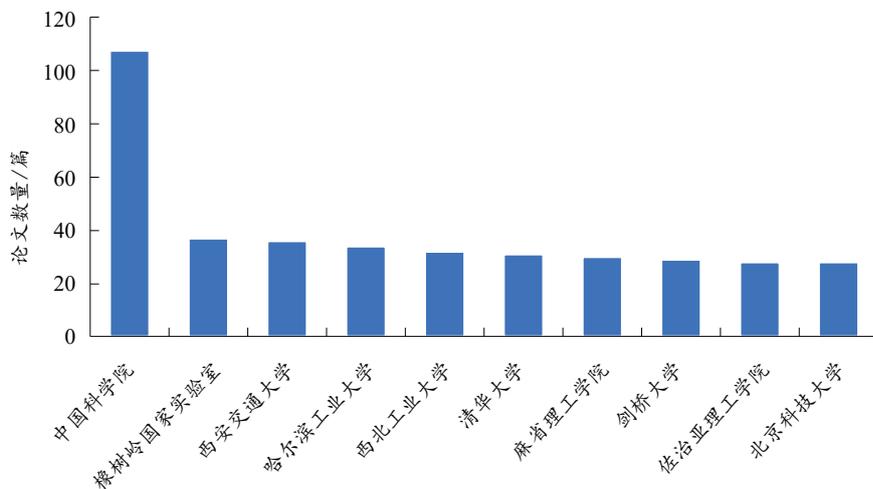


图 3 AI 驱动的材料化学研究论文数量 Top10 机构



表 4 AI 驱动的材料化学研究领域论文数量 Top10 机构合作情况

机构名称	中国科学院	橡树岭国家实验室	西安交通大学	哈尔滨工业大学	西北工业大学	清华大学	麻省理工学院	剑桥大学	佐治亚理工学院	北京科技大学
中国科学院	108	1	1	1	4	3	1	1	3	2
西安交通大学	0	0	36	2	4	0	1	0	0	2

注：表中数据表示机构合作论文数量，单位为篇。

对AI驱动的材料化学研究的高被引论文进行分析。美国发表高被引论文的机构包括美国国家标准与技术研究院^[16-17]、美国西北大学^[18-19]、杨百翰大学^[20]、哈佛大学^[21]、杜克大学^[22]、佐治亚理工学院^[22]、宾西法尼亚州立大学^[23]和美国哈弗福德学院^[24]。欧洲和英国机构包括德国波鸿鲁尔大学^[25]、德国哈勒-维腾贝格大学^[26]、德国马克斯·普朗克可持续材料研究所^[27]、俄罗斯俄罗斯斯科尔科沃科学技术研究院^[28]、莫斯科物理技术学院^[29]以及剑桥大学^[29]。日本主要机构包括日本东北大学^[29]和日本国立材料科学研究所（NIMS）^[30]。中国主要机构包括南开大学^[29-30]、上海大学^[31-32]、西北工业大学^[33]、天津大学^[34]、中国电子科技大学^[35]、浙江大学^[35]和郑州大学^[36]等。

在AI驱动的材料化学研究领域，关键研究机构发文章量和高被引论文分析揭示了全球范围内重要论文产出机构，中国科学院和橡树岭国家实验室等机构发文量位居前列，同时，包括美国国家标准与技术研究院、美国西北大学、南开大学和上海大学等在内的多个国内外机构也在高被引论文发表方面表现较好。

3.4 AI 驱动的材料化学研究重要进展

基于通用科技文本可视化挖掘系统 ITGinsight 对论文的研究主题进行分析，按照论文数量对主题词进行排序（图 4）。可以看出，机器学习、人工神经网络、材料性能、物理性质、随机森林和深度学习等主题词出现频率较高。

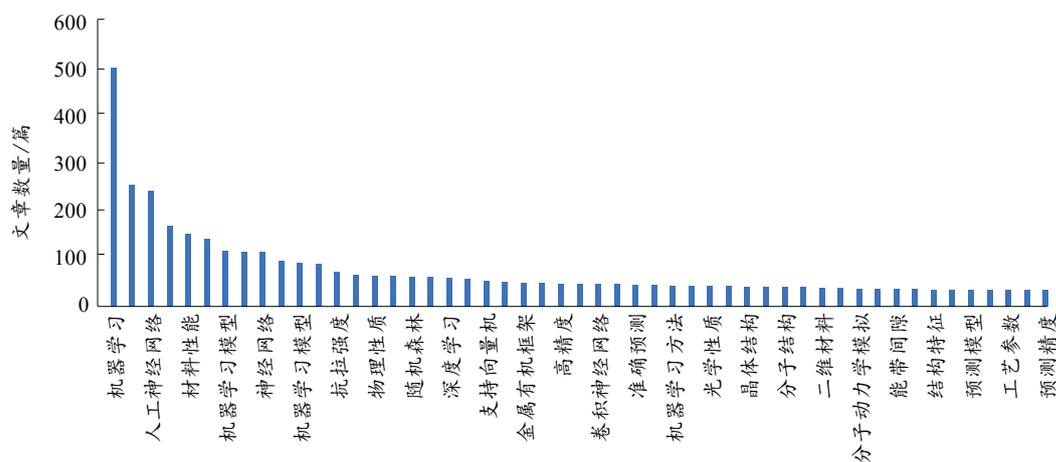


图 4 基于 ITGinsight 得到的高频主题词及其在文章中出现的频次



本部分基于 ITGinsight 以及人工判读, 将主题词分为 5 个类别。

1) 人工智能技术。记录数量较多的主题词包括: 机器学习、人工神经网络、遗传算法、神经网络、随机森林、深度学习、支持向量机、卷积神经网络、回归模型等。

2) 材料。记录数量较多的主题词包括: 新材料、复合材料、二维材料、金属有机框架、高熵合金、功能材料、碳纳米管、纳米多孔材料、多孔材料、金属卤化物钙钛矿等, 材料性能主要包括抗拉强度、抗压强度、电气性能、抗弯强度、弹性特性和拉伸性能等。

3) 材料的结构化学。记录数量较多的主题词包括: 晶体结构、分子结构、电子结构、材料结构、原子结构、结构特性、结构 - 性质关系、物理性质、力学性质、电子性质、光学性质、能带间隙等。

4) 材料制备过程的热力学。主要研究气体和溶液的热力学、吉布斯自由能变化、相图和相变, 记录数量较多的主题词包括: 玻璃化转变温度、相位图、热导率、相变、热性能等。

5) 材料制备过程的动力学。主要研究传质理论、化学反应速率和化学反应的控制步骤, 记录数量较多的主题词包括: 分子动力学、化学性质、化学成分、化学结构、材料成分、热稳定性、合金成分、多相催化、结构稳定性、电化学性质、工艺参数、增材制造等。

基于以上分类, 本部分以主要项目承担机构和高水平论文发表机构为线索, 并根据研究内容相关性进行扩展, 得出 AI 技术、材料的结构化学、材料制备过程的热力学, 以及材料制备过程的动力学 4 个方向重要研究进展, 分析 AI 技术在哪些方面加速材料化学领域的研究进程以及该领域面临的挑战。

3.4.1 人工智能技术全面深入发展, 与领域结合更紧密

AI 技术取得显著进展, 特别是在设计和优化数据库、开发高效材料预测工具包、构建机器学习模型、与领域知识融合, 以及开发通用工具/方法等方面, 例如, 瑞士洛桑联邦理工学院基于机器学习设计并优化了包含 20000 个假设金属有机框架的数据库, 改善数据库的多样性指标, 评估了不同结构在碳捕获和储氢应用中的性能, 证明了多种结构的性能优于基准材料^[37]; 美国纽约

州立大学布法罗分校开发了探索水热合成最有效析氢反应催化剂的机器学习工具包^[38]; 哈尔滨工业大学结合合金的电子和结构特性构建准确有效的机器学习模型, 并预测析氢反应高性能合金催化剂^[39]; 橡树岭国家实验室开发了工作流程和测试平台评估图神经网络 (GNN), 针对材料化学应用提出将领域知识与 GNN 模型结合, 以及通过迁移学习整合现有数据集知识等研发建议^[40]; 美国西北大学提出通用机器学习方法预测无机材料性能, 基于计算神经网络确定精确结构 - 性质关系, 开发通用计算技术, 阐明转变和弛豫温度, 以及聚合物材料的各种其他物理和机械性能^[41]; 应用物理与计算数学研究所与香港城市大学开发通用机器学习势, 对材料性质进行原子建模^[42]; 美国纽约州立大学布法罗分校提出机器学习模型部署、领域适用性和模型持久性的指导方针^[43]; 日本京都大学评估比较 4 种类型回归预测合金熔化温度方法, 找到单组分和二元组分固体最佳预测方法^[44]。

3.4.2 材料结构化学方向: AI 技术解决数据异构以及数据量有限等挑战, 桥接微观结构与性能

该方向重要进展主要包括: 橡树岭国家实验室基于深度神经网络从原子分辨图像提取原子位置和缺陷类型等信息^[45]; 佐治亚理工学院将原子和形态结构特征映射到性质上, 训练机器学习模型快速预测新聚合物性质; 麻省理工学院基于神经网络评估无机复合物自旋态分裂, 使用遗传算法发现非常规自旋交叉复合物^[46]; 剑桥大学基于机器学习模型推荐与已知热电材料具有不同性质的材料^[47]; 中国科学院大学和中国科学院新疆理化技术研究所等机构结合机器学习方法, 打通化学成分、晶体结构到合理合成的途径, 预测 3887 种化学成分带隙、5 种热力学稳定和 50 种亚稳态的新三元硒化物, 合成的 LiGaSe_2 具有稳健的二次谐波生成响应和覆盖 2 个大气窗口的长波红外吸收性能^[48]; 日本国立材料科学研究所使用经过训练的分子设计算法识别结构 - 性能定量关系, 确定数千种潜在聚合物, 合成的聚合物导热系数与非复合热塑性塑料中最先进的聚合物相当^[49]; 西北工业大学基于神经网络对 Ti-6Al-4V 合金显微组织与性能相关性建模^[50], 通过集成学习预测高熵合金的可解释硬度^[51], 将机器学习与特征工程结合寻找大压电常数 BaTiO_3 基陶瓷^[52]。上述进展显示, AI 加快了新材



料研发的速度，展示了处理大量数据、预测材料性质，以及设计全新高性能材料的潜力。

3.4.3 材料制备过程的热力学研究方向：AI 加深了对相变影响因素、相稳定性、热传递等的理解

该方向重要进展主要包括：约翰霍普金斯大学基于深度学习模型从材料 X 射线衍射测量结果中识别特定相^[53]；宾夕法尼亚州立大学帕克分校等机构利用监督机器学习，研究点缺陷扩展结构、动态演化及其在诱导单层 MoS₂ 中半导体和金属相变中的作用^[54]；法国新能源和原子能委员会新能源技术创新实验室使用随机森林预测 half-Heusler 化合物的稳定性，基于机器学习模型得到 481 个稳定材料^[55]；德国马克斯普朗克钢铁研究所结合机器学习，快速、自动发现具有最佳热、磁和电性能的高熵合金^[56]；哈尔滨工业大学开发准确高效的机器学习势研究 3 种金属有机框架中的热传递^[57]；瑞士洛桑联邦理工学院基于集成机器学习模型，探索材料的静态和动态行为，实现了能量、极化和实验行为的准确预测^[58]。上述进展显示，AI 具有从数据中识别材料特定相、研究材料相变、预测化合物稳定性，以及研究材料能量传递等方面的能力。

3.4.4 材料制备过程的动力学方向，AI 技术助力描述、模拟和预测化学反应，以及优化化学反应过程

该方向重要进展主要包括：英国剑桥大学结合机器学习、密度泛函约束和密度泛函理论模拟，模拟四面体无定形碳表面化学活性^[59]；IBM 欧洲研究中心使用预训练语言模型 BERT 预测化学反应分类，分类精度最高可达 98.2%，远高于传统方法（41%）^[60]；英国曼彻斯特大学通过训练深度神经网络模型自动阐明反应机理类别，无需任何额外用户输入，且精度高^[61]；美国内布拉斯加大学林肯分校将多任务学习、多输入和图形注意力网络结合，构建能够基于单体化学结构预测反应率的模型，可与宏观动力学模型结合设计新型共聚物^[62]；上海大学基于机器学习方法预测中间产物在钙钛矿氧化物表面的吸附自由能及其过电势，可用于描述表面化学反应，并加速筛选潜在钙钛矿催化剂^[63]；瑞士洛桑联邦理工学院基于机器学习和领域知识确定金属有机框架中金属阳离子的氧化态，该策略具有鲁棒性，可用于检测结构数据库中的错误标注^[64]；麻省理工学院结合密度泛函理论

和精确神经网络势，研究水溶剂中电催化纳米合金的组成和活性^[65]；英国利物浦大学移动机器人使用贝叶斯搜索算法自动进行 688 次搜索光催化剂实验^[66]；英国格拉斯哥大学通用可编程化学合成机器“Chemputer”可使用 1 个平台架构执行 17 种不同反应^[61]。这些进展显示，AI 具有描述、模拟和预测化学反应，以及优化实验过程等方面的能力。

3.4.5 该领域面临的若干挑战

通过概述本文检索到的综述文章中的观点，本部分列出 AI 驱动的材料化学研究的部分挑战与机遇，主要集中在 AI 与材料化学融合交叉的领域。

1) 材料化学研究需要 AI 辅助预测多种材料性质，AI/机器学习依赖于数据库的质量，包括数据的规模、准确性和一致性等。然而，材料化学的实验物性数据的匮乏、不同测量的不一致性、DFT 自身精度的局限性等，都限制了材料化学数据库的发展。目前多个材料数据库并存，且多以 DFT 计算值为主，亟待整合和扩充成一个全面的“材料大数据库”^[67]。

2) 算法是 AI 和数据驱动的材料化学创新的核心，由于算法基于数学、以及材料科学问题通常有隐藏的物理规律等固有因素，一般算法仍然无法完全满足材料科学的需求。此外，在当前的化学和材料科学研究中，化学和材料科学领域的研究人员倾向于使用现有的算法和模型来解决问题，所产生的效果尚不具有颠覆性。合理设计和选择专门针对材料科学的算法可以帮助科学家更好地理解材料科学问题并节省计算资源^[59]。

3) 描述符是数据驱动过程中人与机器之间传递信息的桥梁，存在数据库描述符不足、描述符不通用、实验描述符和模拟描述符不兼容，以及小样本数据集描述符不可访问等挑战。未来发展机会包括：使用主动学习、机器人辅助的高通量实验和 AI 方法开发新的组合描述符；开发可以从数据库中经济地获取或计算并精确呈现复杂性质的描述符，使得描述符与算法和数据库的发展保持同步；至关重要是，开发方法和理论，推导出表现优异的描述符，并整合多种数据形式^[59]。

4) 虽然理论上可以预测许多低能的亚稳态晶体结构，但只有少数预测结构可能被实际合成，目前还没有明确的、可合成亚稳态结构的遴选规则。未来有望通过



在某些条件下将亚稳态降低到热力学稳定性，或者从动力学的角度从稳定的构建块组装亚稳态相等方法判断亚稳态结构的可合成性^[68]。

5) 目前的数学和数值模型在计算上昂贵，并且通常不能同时捕捉到二维材料生长过程中涉及的所有物理现象，限制了模型在二维材料生长主动控制中的广泛应用。此外，原位实验监测生长技术也无法支持对计算模型预测结果的验证。长期具有挑战性的研发目标是开发计算模型，以及研究其在二维材料设计和合成中的应用，包括但不限于：整合适用于不同尺度的方法，开展二维材料生长的高保真预测；使用二维材料结构和生长条件的实验和计算数据训练机器学习模型^[69]。

6) 电池材料的电化学建模涉及固体扩散、电化学反应、电解质传输、电子传导、化学-力学耦合模型，以及必要时的相分离热力学模型。在锂嵌入材料的基础研究过程中，人们提出了各种多孔电极模型，通过理论分析完善了相分离材料和反应动力学模型，但尚未在这些模型中建立微观结构与宏观性质间的关联。未来，应结合理论、数据驱动方法和微观结构数据集共同推进电化学建模发展^[70]。

综上所述，AI 在材料化学领域的应用面临着多重挑战，包括数据库质量提升、算法适用性增强、描述符优化、预测结果可用性、数学和数值模型改进，以及结合领域知识等。

4 结束语

本文调研了世界主要国家在人工智能驱动的材料化学研究领域部署的项目，检索并分析了该领域的论文，综合项目和论文数据，本文分析了该领域研究主题和重要研究进展，概述了该领域面临的挑战。研究表明，人工智能驱动的材料化学研究是近 10 年来快速发展的新兴领域。人工智能技术全面发展，并与材料化学融合，正在为材料化学研究发展带来前所未有的创新活力，加快推进该领域深入发展，有助于解决复杂问题，尤其是在加快先进材料的发现方面具有重要作用。未来，该领域的发展仍有赖于 AI 与材料化学的深度融合，有待建立高质量的数据库，开发专门针对材料科学问题的、能够揭示隐藏物理规律的先进和节能的算法，以及开发更精准的预测方法、更高效的计算模型。

参考文献

- [1] Nature. Materials chemistry articles from across Nature Portfolio[EB/OL]. [2024-09-01]. <https://www.nature.com/subjects/materials-chemistry>.
- [2] European Commission. Materials design at the exascale. European Centre of Excellence in materials modelling, simulations, and design[EB/OL]. [2024-09-01]. <https://cordis.europa.eu/project/id/824143/reporting>.
- [3] European Commission. A European AI on demand platform and ecosystem[EB/OL]. [2024-09-01]. <https://cordis.europa.eu/project/id/825619/reporting>.
- [4] European Commission. The materials genome in action[EB/OL]. [2024-09-01]. <https://cordis.europa.eu/project/id/666983>.
- [5] European Commission. The novel materials discovery laboratory[EB/OL]. [2024-09-01]. <https://cordis.europa.eu/project/id/676580>.
- [6] European Commission. Novel materials discovery[EB/OL]. [2024-09-01]. <https://cordis.europa.eu/project/id/951786>.
- [7] EPSRC. Integration of computation and experiment for accelerated materials discovery[EB/OL]. [2024-09-01]. <https://gow.epsrc.ukri.org/NGBOViewGrant.aspx?GrantRef=EP/N004884/1>.
- [8] UKRI. £100m boost in AI research will propel transformative innovations[EB/OL]. [2024-09-01]. <https://www.ukri.org/news/100m-boost-in-ai-research-will-propel-transformative-innovations/>.
- [9] NIMS. Materials research by information integraton initiative (MI2I)[EB/OL]. [2024-09-01]. <https://www.nims.go.jp/MII-I/en/>.
- [10] Toyota Research Institute. Materials discovery[EB/OL]. [2024-09-01]. <https://www.tri.global/materials-discovery>.
- [11] Toyota Research Institute. High-throughput polymer discovery (HTP)[EB/OL]. [2024-09-01]. <https://www.tri.global/high-throughput-polymer-discovery-htp>.
- [12] 客观日本. 日本成立新的产官学组织，加速材料开发[EB/OL]. [2024-09-01]. https://www.keguanjp.com/kgjp_jish/kgjp_jish/pt20210322000001.html. (Objective Japan. Japan



- establishes a new industry government university organization to accelerate material development[EB/OL].[2024-09-01].https://www.keguanjp.com/kgjp_jish/kgjp_jish/pt20210322000001.html.)
- [13] 客观日本.日本文部科学省确定2024年度“战略性创造研究推进事业”的战略目标[EB/OL].[2024-09-01].https://www.keguanjp.com/kgjp_zhengc/kgjp_zhengc/pt20240410000003.html.(Objective Japan. Ministry of Education, Culture, Sports, Science and Technology has set the strategic goal of promoting strategic creative research for the year 2024[EB/OL].[2024-09-01].https://www.keguanjp.com/kgjp_zhengc/kgjp_zhengc/pt20240410000003.html.)
- [14] Charles W U, Douglas A S, Bobby G S, et al. Computational neural networks and the rational design of polymeric materials: the next generation polycarbonates[J].*Computational and Theoretical Polymer Science*,1998,8(3/4):311-321.
- [15] Brasquet C, Cloirec P L.Effect of activated carbon cloth surface on organic adsorption in aqueous solutions. use of statistical methods to describe mechanisms[J].*Langmuir*,1999,15(18):5906-5912.
- [16] Choudhary K, DeCost B. Atomistic line graph neural network for improved materials property predictions[J]. *npj Computational Materials*,2021,7(1):185.
- [17] Choudhary K, DeCost B, Chen C, et al. Recent advances and applications of deep learning methods in materials science[J]. *npj Computational Materials*,2022,8(1):59.
- [18] Logan W, Ankit A, Alok C, et al. A general-purpose machine learning framework for predicting properties of inorganic materials[J]. *npj Computational Materials*,2016,2(1):16028.
- [19] Peyman Z M, Yongchul G C, Randall Q S. Progress toward the computational discovery of new metal-organic framework adsorbents for energy applications[J].*Nature Energy*,2024,9(2):121-133.
- [20] Gus L W H, Tim M, Cormac T, et al.Machine learning for alloys[J]. *Nature Reviews Materials*,2021,6(8):730-755.
- [21] Pascal F, Florian H, Jonny P, et al.Machine-learned potentials for next-generation matter simulations[J]. *Nature Materials*,2021,20(6):750-761.
- [22] Chiho K, Anand C, Tran D H, et al.Polymer genome: a data-powered polymer informatics platform for property predictions[J].*The Journal of Physical Chemistry C*, 2018,122:17575-17585.
- [23] Li W, Yang T, Liu C, et al.Optimizing piezoelectric nanocomposites by high-throughput phase-field simulation and machine learning [J].*Advanced Science*, 2022, 9(13):2105550.
- [24] Paul R, Katherine C E, Philip D F A, et al.Machine-learning-assisted materials discovery using failed experiments[J]. *Nature*,2016, 533(7601):73-76.
- [25] Tiago F T C, Antonio S, Miguel A L M.Sampling the materials space for conventional superconducting compounds[J]. *Advanced Materials*,2023, 36(1):2307085.
- [26] Jonathan S, Mário R G M, Silvana B, et al.Recent advances and applications of machine learning in solid-state materials science[J].*npj Computational Materials* ,2019, 5:83.
- [27] Rao Z Y, Yen P, Xie R W, et al.Machine learning-enabled high-entropy alloy discovery[J].*Science*,2022,378(6615):78-85.
- [28] Artem R O, Chris J P, Zhu Q, et al.Structure prediction drives materials discovery[J].*Nature Reviews Materials*,2019, 4(5): 331-348.
- [29] Wu S, Kondo Y, Kakimoto M, et al.Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm[J].*npj Computational Materials*,2019,5(1):66.
- [30] Liu J Y, Liu H, Chen H J, et al.Progress and challenges toward the rational design of oxygen electrocatalysts based on a descriptor approach[J].*Advanced Science*,2019,7(1):1901614.
- [31] Liu Y, Guo B, Zou X X, et al.Machine learning assisted materials design and discovery for rechargeable batteries[J]. *Energy Storage Materials*,2020,31: 434-450.
- [32] Liu Y, Zhao T L, Ju W W, et al.Materials discovery and design using machine learning[J].*Journal of Materiomics*,2017,3(3):159-177.
- [33] Oganov A R, Pickard C J, Zhu Q, et al.Structure prediction drives materials discovery[J].*Nature Reviews Materials*,2019,4(5):331-348.
- [34] Artem R O, Chris J P, Zhu Q, et al.Theory-guided design of catalytic materials using scaling relationships and reactivity descriptors[J].*Nature Reviews Materials*,2019,4(12):792-804.
- [35] Li W X, Yang T N, Liu C S, et al.Optimizing piezoelectric nanocomposites by high-throughput phase-field simulation and machine learning[J].*Advanced Science*,2022,9(13):2105550.
- [36] Zhang X, Chen A , Chen L, et al.2D Materials bridging experiments and computations for electro/photocatalysis[J]. *Advanced Energy Materials*,2021,12(4):2003841.
- [37] Sauradeep M, Seyed M, Moosavi K, et al.Diversifying databases of metal organic frameworks for high-throughput computational screening[J].*ACS Applied Materials & Interfaces*,2021,13(51):61004-61014.
- [38] Wei S,Baek S,Yue H, et al.Machine-learning assisted exploration: toward the next-generation catalyst for hydrogen evolution reaction[J].*Journal of the Electrochemical Society*,



2021, 168:12.

[39] Zhang J Z, Wang Y L, Zhou X Y, et al. Accurate and efficient machine learning models for predicting hydrogen evolution reaction catalysts based on structural and electronic feature engineering in alloys[J]. *Nanoscale*, 2023, 15(26):11072-11082.

[40] Victor F, Zhang J X, Eric J, et al. Benchmarking graph neural networks for materials chemistry[J]. *npj Computational Materials*, 2021, 7(1):84.

[41] Logan W, Ankit A, Alok C, et al. A general-purpose machine learning framework for predicting properties of inorganic materials[J]. *npj Computational Materials*, 2016, 2(1):16028.

[42] Wen T, Wang R, Zhu L, et al. Specialising neural network potentials for accurate properties and application to the mechanical response of titanium[J]. *npj Computational Materials*, 2021, 7(1):206.

[43] Gaurav V, Doaa A, Ramachandran S, et al. ChemML: a machine learning and informatics program package for the analysis, mining, and modeling of chemical and materials data [J]. *WIREs Computational Molecular Science*, 2020, 10(4):e1458.

[44] Seko A, Maekawa T, Tsuda K, et al. Machine learning with systematic density-functional theory calculations: application to melting temperatures of single and binary component solids[J]. *Physical Review B*, 2014, 89:054303.

[45] Maxim Z, Orlin D, Artem M, et al. Deep learning of atomically resolved scanning transmission electron microscopy images: chemical identification and tracking local transformations[J]. *ACS Nano*, 2017, 11(12):12742-12752.

[46] Janet J P, Chan L, Kulik H J. Accelerating chemical discovery with machine learning: simulated evolution of spin crossover complexes with an artificial neural network[J]. *Journal of Physical Chemistry Letters*, 2018, 9(5):1064-1071.

[47] Gaultois M W, Olynyk A O, Mar A, et al. Perspective: web-based machine learning models for real-time screening of thermoelectric materials properties[J]. *APL Materials*, 2016, 4(5):199-205.

[48] Cai W B, Ailijiang A, Xi C W, et al. Toward the rational design of mid-infrared nonlinear optical materials with targeted properties via a multi-level data-driven approach[J]. *Advanced Functional Materials*, 2022, 32(23):2200231.

[49] Wu S, Kondo Y, Kakimoto M, et al. Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm[J]. *npj Computational Materials*, 2019, 5(1):66.

[50] Sun Y, Zeng W, Han Y, et al. Modeling the correlation between microstructure and the properties of the Ti-6Al-4V alloy

based on an artificial neural network[J]. *Materials Science and Engineering: A*, 2011, 528(29/30):8757-8764.

[51] Zhang Y F, Ren W, Wang W L, et al. Interpretable hardness prediction of high-entropy alloys through ensemble learning[J]. *Journal of Alloys and Compounds*, 2023, 945:169329.

[52] Yuan R, Yuan R, Xue D, et al. Machine learning combined with feature engineering to search for BaTiO₃ based ceramics with large piezoelectric constant[J]. *Journal of Alloys and Compounds*, 2022, 908:164468.

[53] Nam Q L, Michael P, Alexander N, et al. Deep learning models to identify common phases across material systems from x-ray diffraction[J]. *The Journal of Physical Chemistry C*, 2023, 127(44):21758-21767.

[54] Tarak K P, Zhang F, Daniel S S, et al. Defect dynamics in 2-D MoS₂ probed by using machine learning, atomistic simulations, and high-resolution microscopy[J]. *ACS Nano*, 2018, 12(8):8006-8016.

[55] Legrain F, Carrete J, Van R A, et al. Materials screening for the discovery of new half-Heuslers: machine learning versus ab initio methods[J]. *Journal of Physical Chemistry B*, 2017, 122(2):625-632.

[56] Hu Q, Yang R. The endless search for better alloys[J]. *Science*, 2022, 378:26-27.

[57] Ying P H, Liang T, Xu K, et al. Sub-micrometer phonon mean free paths in metal-organic frameworks revealed by machine-learning molecular dynamics simulations[J]. *ACS Applied Materials and Interfaces*, 2023, 15(30):36412-36422.

[58] Gigli L, Veit M, Kotiuga M, et al. Thermodynamics and dielectric response of BaTiO₃ by data-driven modeling[J]. *npj Computational Materials*, 2022, 8(1):209.

[59] Volker L D, Miguel A C, Richard J, et al. Computational surface chemistry of tetrahedral amorphous carbon by combining machine learning and density functional theory [J]. *Chemistry of Materials*, 2018, 30(21):7438-7445.

[60] Philippe S, Daniel P, Alain C V, et al. Mapping the space of chemical reactions using attention-based neural networks [J]. *Nature Machine Intelligence*, 2021, 3(2):144-152.

[61] Jordi B, Igor L. Organic reaction mechanism classification using machine learning [J]. *Nature*, 2023, 613(7945):689-695.

[62] Tung N, Mona B. Machine learning approach to polymer reaction engineering: determining monomers reactivity ratios[J]. *Polymer*, 2023, 275:125866.

[63] Wang X M, Xiao B, Li Y H, et al. First-principles based machine learning study of oxygen evolution reactions of perovskite oxides using a surface center-environment feature



- model[J].Applied Surface Science,2020,531:147323.
- [64] Jablonka K M, Ongari D, Moosavi S M, et al. Using collective knowledge to assign oxidation states of metal cations in metal-organic frameworks[J].Nature Chemistry,2021,13(8):771-777.
- [65] Artrith N, Kolpak A M.Understanding the composition and activity of electrocatalytic nanoalloys in aqueous solvents: a combination of DFT and accurate neural network potentials[J]. Nano Letters, 2014, 14(5):2670.
- [66] Benjamin B, Phillip M M, Vladimir V G, et al.A mobile robotic chemist [J].Nature,2020, 583(7815): 237-241.
- [67] Wang Z, Sun Z H, Yin H, et al.Data-driven materials innovation and applications[J].Advanced Materials,2022, 34(36):2104113.
- [68] Artem R O, Chris J P, Zhu Q, et al.Structure prediction drives materials discovery [J].Nature Reviews Materials,2019, 4(5):331-348.
- [69] Kasra M, Ji Y Z, Wang Y X, et al.Multiscale computational understanding and growth of 2D materials: a review [J].npj Computational Materials,2020, 6(1): 22.
- [70] Xu H Y, Zhu J, Donal P F, et al.Guiding the design of heterogeneous electrode microstructures for Li-ion batteries: microscopic imaging, predictive modeling, and machine learning [J].Advanced Energy Materials,2021,11(19):2003908.

Analysis of the International Development Trend of Materials Chemistry Research Driven by Artificial Intelligence

Lü Fengxian^{1*}, Liu Xiaoping^{1,2}, Tao Zhiyu^{1,2}

1. National Science Library, Chinese Academy of Sciences, Beijing 100190, China

2. Department of Information Resources Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China

*Corresponding Author, E-mail:lyufx@mail.las.ac.cn

[Abstract] [Objective/Significance] The paradigm of scientific research driven by artificial intelligence has undergone a transformation, and materials chemistry is one of its important fields. [Method/Process] This article focuses on the research of artificial intelligence driven materials chemistry field, investigates the project deployment of major countries in this field, and combines the papers in this field in the WoS database to comprehensively analyze the main countries, important publishing institutions, important research progress, and challenges faced in the field using analytical tools such as DDA and ITGinsight. [Results/Conclusions] Research has found that material chemistry research driven by artificial intelligence has received attention from major countries such as the United States, Europe, the United Kingdom, and Japan, and is a rapidly developing emerging field in recent years. Artificial intelligence is assisting in accelerating the structural chemistry research of materials, thermodynamic research of material preparation processes, and dynamic research of material preparation processes. Artificial intelligence technology is bringing unprecedented innovative vitality to material chemistry research and accelerating the in-depth development of this field. It helps to solve complex problems, especially in accelerating the discovery of advanced materials. In the future, AI driven material chemistry research still relies on the deep integration of artificial intelligence and material chemistry.

[Keywords] artificial intelligence, materials chemistry, important research progress, challenge



附录

论文检索式

(TS=("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network" OR "reinforcement learning" OR "genetic algorithm" OR "support vector machine" OR "decision tree" OR "random forest" OR "unsupervised learning") and Ts(("Structur* Chemistry" and Material*) or (((composit* or structur*) and propert*) and material*) or (Thermodynamic* and "material prepar*") or ((gas* or solution*) and (Thermodynamic* or "Gibbs free energy" or "phase diagram*" or "phase transition*")) or ((Kinetic* or Dynamic*) and "material prepar*") or "Mass transfer theory" or ("Chemical reaction" and material*) or ("chemical prepar*" and material*) or (("material* chemistry" or ((composit* or structure*) and propert*)) and (Test* or characteriz*))) And so=((nature chemistry) or (nature materials) or (Nature Reviews Materials) or (nature reviews chemistry) or (nature) or (science) or (Nature Computational Science))) or ((wc=(COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE) or TS=("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network" OR "reinforcement learning" OR "genetic algorithm" OR "support vector machine" OR "decision tree" OR "random forest" OR "unsupervised learning")) and Ts(("Structur* Chemistry" and Material*) or (((composit* or structur*) and propert*) and material*) or (Thermodynamic* and "material prepar*") or ((gas* or solution*) and (Thermodynamic* or "Gibbs free energy" or "phase diagram*" or "phase transition*")) or ((Kinetic* or Dynamic*) and "material prepar*") or "Mass transfer theory" or ("Chemical reaction" and material*) or ("chemical prepar*" and material*) or (("material* chemistry" or ((composit* or structure*) and propert*)) and (Test* or characteriz*))) and WC=((CELL & TISSUE ENGINEERING) OR (CONSTRUCTION & BUILDING TECHNOLOGY) OR (CRYSTALLOGRAPHY) OR (ENGINEERING, ELECTRICAL & ELECTRONIC) OR (ENGINEERING, MANUFACTURING) OR (MATERIALS SCIENCE, BIOMATERIALS) OR (MATERIALS SCIENCE, CERAMICS) OR (MATERIALS SCIENCE, CHARACTERIZATION & TESTING) OR (MATERIALS SCIENCE, COATINGS & FILMS) OR (MATERIALS SCIENCE, COMPOSITES) OR (MATERIALS SCIENCE, MULTIDISCIPLINARY) OR (MATERIALS SCIENCE, PAPER & WOOD) OR (MATERIALS SCIENCE, TEXTILES) OR (METALLURGY & METALLURGICAL ENGINEERING) OR (NANOSCIENCE & NANOTECHNOLOGY) OR (PHYSICS, CONDENSED MATTER) OR (POLYMER SCIENCE)) and WC=((BIOCHEMICAL RESEARCH METHODS) OR (BIOCHEMISTRY & MOLECULAR BIOLOGY) or (CHEMISTRY, ANALYTICAL) or (CHEMISTRY, APPLIED) or (CHEMISTRY, INORGANIC & NUCLEAR) or (CHEMISTRY, MEDICINAL) or (CHEMISTRY, ORGANIC) or (CHEMISTRY, PHYSICAL) or (ELECTROCHEMISTRY) or (ENERGY & FUELS) or (ENGINEERING, CHEMICAL) or (ENGINEERING, PETROLEUM) or (FOOD SCIENCE & TECHNOLOGY) or (GEOCHEMISTRY & GEOPHYSICS) or (GEOLOGY) or (MINERALOGY) or (NANOSCIENCE & NANOTECHNOLOGY) or (OCEANOGRAPHY) or (PHARMACOLOGY & PHARMACY) or (POLYMER SCIENCE) or (SPECTROSCOPY))).