



个人简历

P E R S O N A L R E S U M E



姓名：王丁灏



电话：19311009310



邮箱：1604512149@qq.com



微信：wdh201309296495



QQ号：1604512149



王丁灏 3025004471

住址：北洋园三号院29斋524

志愿：宣怀学院科创营



193-1100-9310



1604512149@qq.com



1604512149

基本信息 | Basic Information

年龄：18周岁

学院：精密仪器与光电子工程学院

年级：大学一年级

班级：工程科学实验班一班（王守融一班）

个人能力 | Personal Skills



个人爱好 | Hobbies



音乐



乒乓



功夫



诗词

自我评价 | Self assessment

我是一名具有卓越能力和不凡潜力的优秀大一新生。无论是在过去的创新创业经历中，还是在未来的大学实践活动中，我的经验、思想和才华都会为团队注入极大活力。在科创营和未来的学习生活中，我都将用大脑严谨思考，用双手躬身实践，为成功铺出一条踏踏实实的坦途。

教育经历 | Education

2013.9 – 2019.7 太原市新建路小学校（总校区）2013级9班

2019.9 – 2022.7 太原市五育中学校 296班

2022.9 – 2025.6 太原市第五中学校 495班

2025.9 – 今 天津大学精密仪器与光电子工程学院 王守融1班

创新创业经验 | Innovative and Entrepreneurship Experience

1 数学建模大赛团体一等奖 2022.11

身份：组员

奖项级别：校级

在本次活动中：我率先提出《探究手机电量与使用时长之间的关系》；独立完成小米11青春版手机的纵向实验测试；参与班级演讲，争取到了年级评比机会。

2 《锰彩风华》课题研究展示 2024.5

身份：化学社社长、课题组长 展示级别：校级

在本次活动中：我主导了整个课题的研究脉络，敲定展示流程，与社员一道，直接参与学校“双新”展示。该展示受到北京师范大学裴娣娜教授高度评价：“这是我在全国见到的，学生自己发现问题、提出问题、解决问题的第一例。”

3 线上和家庭教师 2022.7 - 今

身份：教师

中考后，在B站发表视频，讲解初中数学竞赛和高中新学的知识，现拥有2700+粉丝。

高考后，担任两名学生的家庭教师。现已帮助一名学生在分班考中考入其所在学校的最好的班型。

座右铭 | Motto

男儿到死心如铁，看试手，补天裂。

——辛弃疾

我是宝剑，我是火花，我愿生如闪电之耀亮，我愿死如彗星之迅忽。

——高君宇

自荐信

尊敬的宣怀学院的各位老师：

您好！

我是来自山西太原的王丁灏，毕业于太原五中，现就读于天津大学精密仪器与光电子工程学院工程科学实验班1班（王守融1班）。非常荣幸能够将我的个人简历投递在您面前。现在，我将在这封自荐信中向各位老师进行自我推荐。

经过高考的测验，自高中就秉持创业梦的我有幸考入天津大学这座百卅学府。在这里，我的师友导师张孟伦老师告诉我，如果想成为一个成功的创业者，宣怀学院将是我的不二选择。于是，我怀着崇敬的心情，前来报名参加宣怀学院科创营。

在过往的岁月中，我担任过很多职务，包括但不限于：班长、体育委员、社团社长、课代表（详见下表“我的干部生涯”）。在这些职务上的工作让我成为了一名类似于“技术型CEO”的全能人才，也为我的创业之路打下了基石。这让我不仅具有宏观把握大方向的能力，也让我可以洞察细微，在每一个细节都可以精雕细琢、严格规范。我相信这些品质正是宣怀学院所需要的，也是我引以为傲的。

10月1日，我有幸在场地启动仪式上与院长刘毅老师交流。在交流中，我深深地感受到：宣怀学院是一个开放、包容的学院，是一个共建、共享的集体。我非常希望成为这个集体的一分子，并在宣怀学院活动的参与过程中，获得创业的知识与能力。就目前我的规划，未来我可能在高新科技或新型教培上开创一方天地，将“创业”扎根在祖国的大地上，为国家经济社会发展贡献出自己的力量。如果创业成功，我将回到山西，发展家乡，让创业成果为父老乡亲谋幸福！

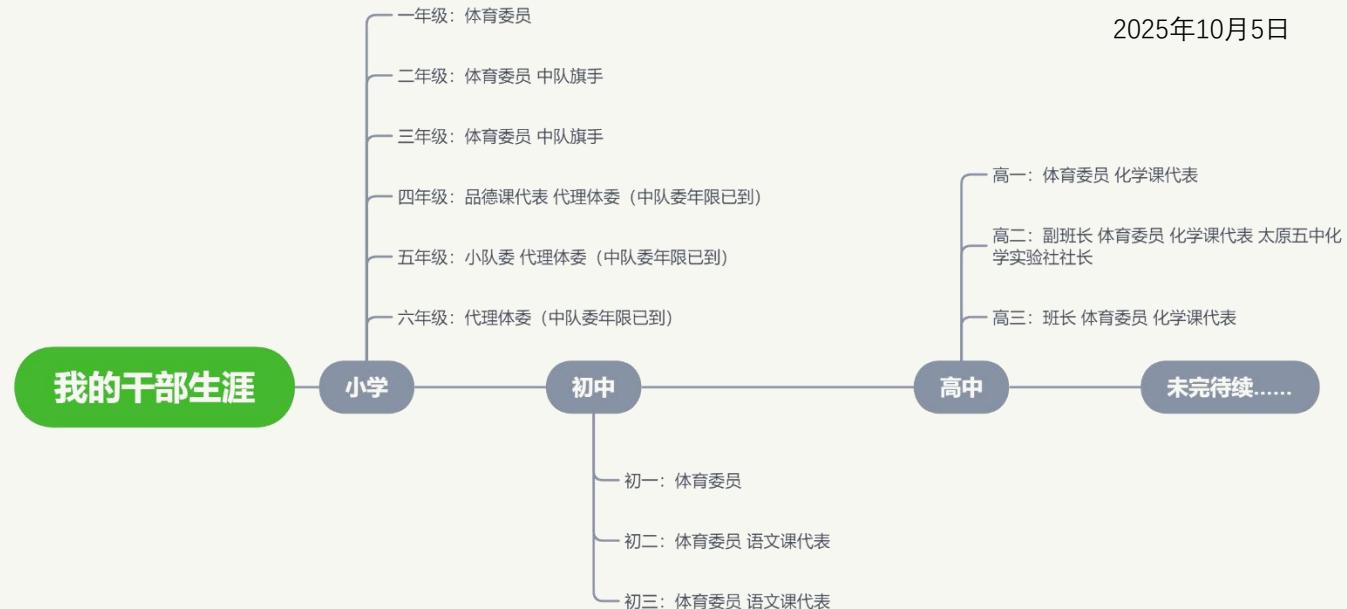
在此，我郑重向各位老师申请加入宣怀学院科创营，成为宣怀学院的一员！

此致

敬礼！

王丁灏

2025年10月5日





期待您的回复

天津大学“未来 30 年”颠覆性创新创想大赛方案申报表



方案名称: “以算解化”——基于理化数据推演的人工智能
化学平台

负责人: 王丁灏

天津大学

2025 年 9 月

基本信息

方案基本信息

方案名称	“以算解化”——基于理化数据推演的人工智能化学平台		
组队形式	<input type="checkbox"/> 个人项目 <input checked="" type="checkbox"/> 团队项目 团队人数: <u>8</u> 人(其中: 指导老师: <u>1</u> 人, 学生: <u>7</u> 人)		
涉及领域 (可多填)	化学、物理学、计算机科学与工程、人工智能、材料学、生物医药、化学工程与技术		

方案简介（200字以内）

利用人工智能可推理的能力和化学“结构决定性质”的基本逻辑，搭建基于物理、化学结构和理化变化数据的算法，使人工智能借助这些数据，拥有可以推算物质结构和理化性质的能力，即利用已知结构的人工智能推算物质性质。研制成功后，将研究成果广泛应用于化学、生物学、材料学、医学、药学等多个领域，拓宽人类现有知识框架，逆向促进理学发展创新和理论体系完善，推动人类社会发展与变革。

第一负责人基本信息

姓名	王丁灏	性别	男	出生年月	2006年10月		
学院/单位	精密仪器与光电子工程学院			专业	精仪与光电信息类		
身份信息	<input type="checkbox"/> 教职工	职称		职务	工号		
	<input checked="" type="checkbox"/> 学生	学历	本科	年级	2025	学号	3025004471
	<input type="checkbox"/> 校外人员	职务		身份证号			
联系电话	19311009310		邮箱	1602512149@qq.com			

通讯地址	天津市津南区雅观路135号三间园29斋					
所在单位意见	申报人是否是本单位员工/学生并同意参加比赛。 <input type="checkbox"/> 是 <input type="checkbox"/> 否 负责人签章： 年 月 日					
团队成员基本信息						
指导教师基本信息						
姓名	性别	所在单位	职称	联系方式	研究领域	
王鹏	男	精密仪器与光电子工程学院	教授	wang_peng@tju.edu.cn	机器视觉几何量测量 空间位姿测量 微振动测试分析及实时监控	
学生基本信息						
姓名	性别	专业	学号	联系电话	微信号/QQ	分工
王丁灏	男	精仪与光电信息类	30250044 71	19311009310	wdh201309296495/1604512149	底层逻辑设计；人员招募；研究项目的开展背景及其必要性；填写申报表
孙翀然	男	精仪与光电信息类	30250044 54	15922003216	Sun342452	整理研发段参考资料；研发工作建立基础数据库；硬件设施规划；研究导入规律

尹志轩	男	精仪与光 电信息类	30250044 13	18526771652	a18526771652	整理研发段 参考资料； 研发工作建 立基础数据 库；研究导 入规律
张颖玥	女	化学拔尖 班	30250021 23	15660892907	2673882155	评估平台实 际应用能 力；检验算 法逻辑；核 验方案可行 性
马尔哈 巴·加勒 哈斯拜 克	男	预科班	70251050 43	15894761939	Ma15894761939/3202715434	设计新方向 的提供；完 整逻辑的设 计与梳理； 算法逻辑的 制作与简 化；数据要 素的盘点； 应用端专门 化研究
陶庆治	男	医学院	30250066 11	15192039006	15192039006	对该在现实 中的应用策 略进行研 究；优化应 用端的设 计；将平台 与其它应用

						端技术结合
赵海兮	男	医学院	30250061 21	13722538667	Zhaohaixi1470963/3988474001	应用方面数据的检索和排版；生活中实际应用的分析；产品影响力评估
校外人员基本信息						
姓名	性别	身份证号	单位	职务	联系方式	分工

项目构想（参考框架）

一、必要性与颠覆性

1.1 项目定义

基于理化数据推演的人工智能化学平台（Artificial Intelligence Chemistry Platform Based on Data，简称 AIChem 平台），是指利用具有高学习和逻辑推理能力的人工智能体，搭建基于物理学、化学原子级及以下数据与理化推演逻辑的系统性平台，从而在化学相关产业上实现从研发到生产的全流程 AI 赋能，推动人类理论化学和以化学为基础的后端工程科学的发展。

1.2 项目开展背景

当前人工智能化学已进入“模型深耕+场景落地”的关键阶段，依托化学大语言模型与实验自动化技术，正逐步破解传统化学研究的效率瓶颈，相关进展均有明确文献支撑。

技术层面，化学大语言模型实现核心突破。上海交大团队研发的 Chemma 模型，将化学反应转化为自然语言任务，在 USPTO-50k 数据集单步逆合成任务中达成 72.2% 的 Top-1 准确率，远超此前 57.7% 的最优结果。该模型无需依赖量子计算，仅通过海量反应数据预训练，即可完成分子设计、产率预测等全流程任务，在 Suzuki-Miyaura 反应中仅用 15 次实验就实现 67% 产率，较传统方法减少 70% 试错量。华东师大发布的 ChemGPT 2.0 模型参数量达 720 亿，基于 378 万条逆合成数据微调，在化学学科能力测试中全面领先 GPT-4，还能精准预测复杂 3D 结构分子的反应物^[1]。基础工具创新同样关键，中科院上海药物所的 ReactSeq 反应描述语言，通过编码分子编辑操作，解决了传统模型无法描述反应动态过程的问题，提升了逆合成预测的可解释性。^[2]此外，神经网络技术已能突破分子体系势能面的“指数墙”困境，实现 11 个原子的丙烷分子振动能量精确求解。

落地应用呈现“人机协作”特征。华东师大同步推出的机器人化学家“华 24”，掌握 370 余条实验动作，结合自动化合成平台可实现农药分子全连续合成，效率提升 80%。Chemma 模型构建的“Co-Chemist”人机协作框架，在未报道的 N 杂环交叉偶联反应中，通过实验反馈微调模型，第二轮即精准推荐高效配体。这类技术已开始重构研究范式：从“经验试错”转向“数据推演”，从“封闭反应空间”探索转向“开放体系自主优化”。

行业仍面临明确挑战。数据层面，实验数据误差、不完整及标准化程度低的问题突出，制约模型泛化能力。模型层面，“黑盒效应”尚未完全破解，尽管 ReactSeq 通过分子编辑操作编码增强了可解释性，但多数模型仍缺乏化学机理层面的清晰输出。此外，物理化学先验知识与算法融合不足，高能量激发态等复杂场景的预测精度仍需提升。

1.3 潜在价值

作为集数据存储、模型架构、逻辑推理、自我纠错于一体的新型战略性人工智能化学平台，AIChem 系统平台系统将为物理和化学乃至其下游相关的全部学科和产业赋能，为人类对于自然科学的理解和对美好生活的建设做出应有的贡献。

1.3.1 完善化学理论

该平台形成成熟体系后，可以仅凭借基本数据信息，推断得到反应相关产物和产率。通过人工智能化学平台的推演，科技工作者可以将得到的计算数据与真实实验数据对比，利用计算与实际实验的误差推理构建新的化学理论，推翻旧的化学理论。

1.3.2 未知分子建模

在科学的研究中，常常会遇到新型的分子。在测定其实验式和分子式后，可以通过检测过程中得到的各类数据与相关数据库中的数据进行比对，利用算法给出合理的分子结构并推断其性质。

这种建模方式实际是现有建模方法的改进（如 ChemSketch），是在仅有少量数据或因客观原因无法得到大量、准确数字时的替代方案（如不稳定元素和极端条件）。这样的建模方式，可以构建在特定的已完成数学建模的影响因子条件下，分子数据的差异，从而进一步推动不同条件下相关产业的研究。

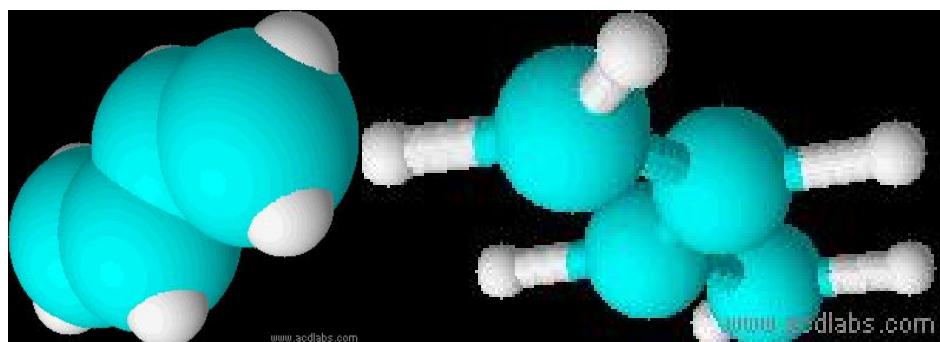


图 1：利用 ChemSketch 完成的 1,3-丁二烯分子建模

1.3.3 新型合成路径开发

AIChem 平台可以对未知基团、物质乃至化学反应的存在性、稳定性、应用性进行评估，为化学合成提供新的研究思路，为人类提供新的合成范式。

1.3.4 新型材料和医药研发

AIChem 平台可以针对特定靶位，自主设计研发具有相应效益的药物，并整体考虑人体健康，尽可能减小药物对人体其它部位的毒害作用。

1.4 国内相关研究基础

1.4.1 人工智能为化学教育赋能

将以知识传授为主要任务的课堂教学交给 AI 完成，人类教师的作用可以主要体现在对学生专业

知识学习的引导、督促、监管，对学生成长过程的辅导、陪伴，以及对学生自主学习能力的培养。此外，人类教师还要及时总结、推广好的学习方式，发现依靠AI组织教学带来的新问题，并加以解决。相信，AI赋能下，化学教育、化学教学，起码是化学专业的理论教学完全可以上一个大的台阶^[3]。

1.4.2 化学家系统“小来”

中国科学技术大学研发的机器化学家系统“小来”，可以完成文献读取、合成、表征、性能测试、机器学习模型建立和优化等全流程任务。中国科学技术大学教授罗毅介绍，通过运用“小来”系统，团队加速了新材料的发现过程，2个月内就完成了需要验证2000年才能完成的复杂优化工作，利用火星陨石制备出实用的产氧电催化剂^[4]。

1.4.3 智能手套箱合成工作站的应用

中国科学院院士、复旦大学教授麻生明团队于2024年6月应用智能手套箱合成工作站，实验效率由此翻了数十倍。不过当下对实验结果的解读还无法依赖机器人。在麻生明看来，目前AI在化学领域的价值仍以提升效率为主，而且必须要注意的是，化学发现很多时候是基于偶然性，但通过已有知识训练出来的机器却常常会泯灭这种偶然“灵光一现”。

1.4.4 人工智能参与的单步逆合成分析过程

北京大学的来鲁华和裴剑锋团队使用了人工智能自然语言处理技术，基于Transformer框架，开发了一种无模板的单步逆合成分析方法。该方法不需要依赖化学反应规则，而是通过输入目标产物分子的SMILES表达式，预测出产物到前体的单步逆合成路径。同时，结合了蒙特卡洛树搜索（MCTS）和启发式打分函数，构建了一个名为AutoSynRoute的逆合成路线规划系统，成功实现了4种分子的逆合成路线设计^[5]。

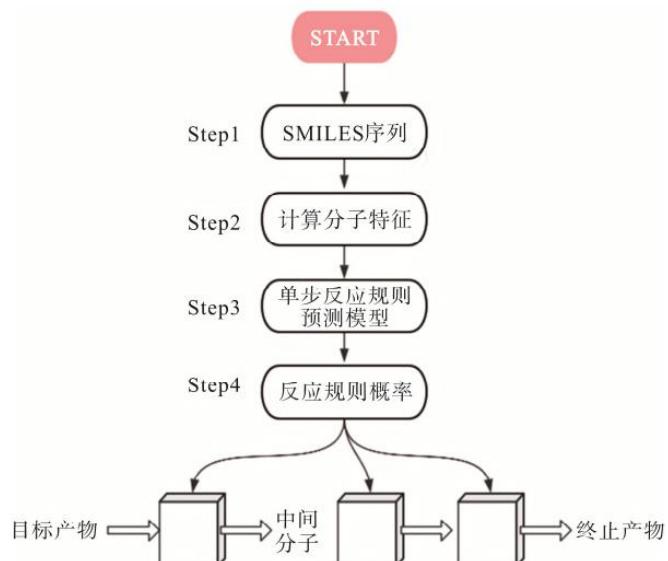


图2：一种单步反应规则预测的生物逆合成方法及系统^[5]

1.4.5 多尺度 AI 框架推动多相催化剂智能设计

传统催化剂研发依赖“试错”，耗时费力。清华团队开发了一套 AI 工具链：

SurFF 模型：预测晶体表面稳定性，效率比传统计算快 10 万倍，筛选出万余种合金表面^[6]。

CO₂电还原催化剂：用 AI 从 3000 种材料中找到高性能单原子催化剂，产甲醇效率提升数倍^[7]。

CaTS 框架：加速过渡态搜索，将反应动力学计算效率提高近万倍。这些工具帮助科学家在原子尺度精准设计催化剂，推动氢能、化工等领域绿色转型^[8]。

1.4.6 贝叶斯学习辅助低铱电解水催化剂设计

电解水制氢需要贵金属铱作催化剂，但成本高昂。清华团队结合 AI 与实验，开发出新型催化剂，利用贝叶斯优化算法，通过模拟筛选出铱掺杂的二氧化钛（Ir-TiO₂），减少铱用量的同时提升活性。新催化剂过电位比商用产品降低 115 毫伏，质量活性提高 23 倍，大幅降低制氢成本。这项技术为大规模氢能生产提供了经济可行的方案^[9]。

1.5 国外相关研究基础

1.5.1 机器人能用 AI 加速发现化学分子

荷兰阿姆斯特丹大学化学家开发了一种自主化学合成机器人。这款被称为“化学机器人”（RoboChem）的台式设备具有集成的、人工智能（AI）驱动的机器学习单元，在速度和准确性方面均胜过人类化学家，同时还表现出高度的独创性。作为同类产品中的第一个，它可显著地加速化学分子发现，用于制药及其他诸多领域^[10]。

1.5.2 全球首个 AI 化学机器人

现在化学实验方法正在变化，越来越多机器人自动操作化学合成平台正在真实科研场景中发生。2024 年 7 月，全球首个 AI 化学机器人登上《自然》杂志封面，它每天工作 21.5 小时，其余时间用来充电，仅 8 天就自主完成 688 个实验，并从中发现了比初始配方活性高 6 倍的光催化剂^[11]，大大加速的新型催化剂发现进程。

1.5.3 在 2024 年颁发的诺贝尔物理学和化学奖与 AI 有关

2024 年诺贝尔化学奖被拆分授予：一半授予戴维·贝克（David Baker），以表彰其在“计算蛋白质设计”领域的贡献；另一半则由德米斯·哈萨比斯（Demis Hassabis）和约翰·江珀（John Jumper）共同获得，以表彰他们在“蛋白质结构预测”领域的成就^[12]。

哈萨比斯提出了“阿尔法折叠”（AlphaFold），该技术能以惊人的精度预测蛋白质的三维结构，这一成果使其获得了 2024 年诺贝尔化学奖。哈萨比斯表示：“我真正兴奋的是将这类人工智能用于科学领域，并推动科学更快发展。几个月前我在欧洲核子研究组织（CERN）发表过一次演讲。我认

为，若有朝一日人工智能能参与发现新粒子，那将非常了不起^[13]。”

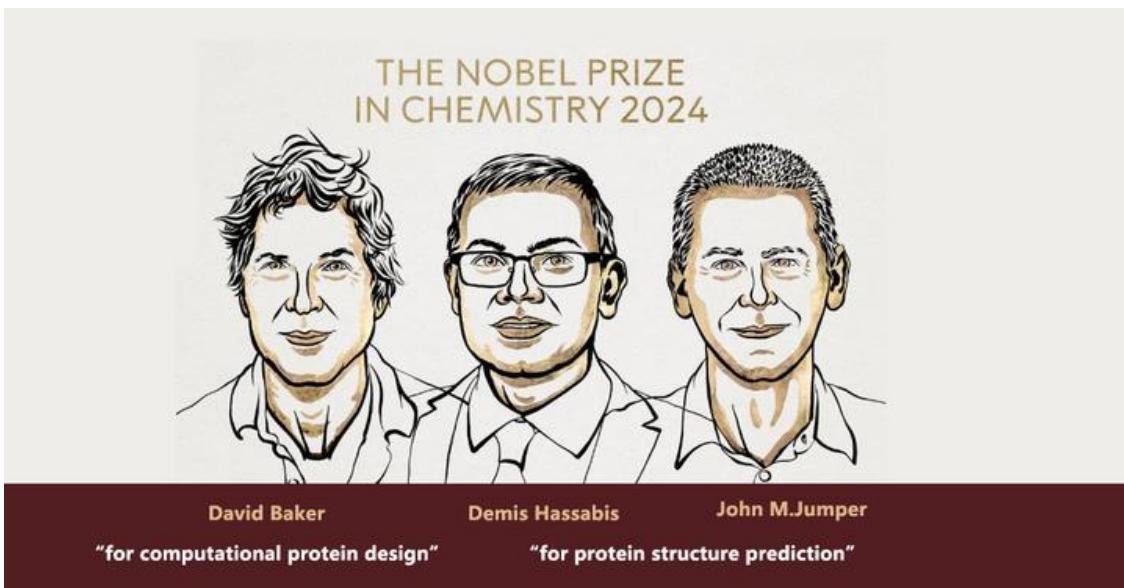


图 3:2024 年诺贝尔化学奖三位得主^[13]

1.5.4 量子力学与机器学习结合预测分子性质

洛斯阿拉莫斯国家实验室构建了量子力学与机器学习融合的预测模型，创新采用半经验量子力学框架结合神经网络。团队用 HIPNN 神经网络替代传统静态参数，动态生成哈密顿量参数，这些参数与原子杂化、键合特性强相关。模型经分子能量、原子力等多任务训练后，兼具可解释性与可迁移性，能精准计算分子激发态动力学，为光—物质相互作用等领域的研究提供高效工具^[14]。

1.5.5 AI 设计高效光催化剂

MIT 借助机器学习完成抗撕裂材料的高通量发现，聚焦 5000 种已合成的二茂铁分子展开筛选。团队先通过力场计算量化 400 种分子的键断裂作用力，再训练模型预测剩余分子的机械响应阈值，最终锁定 m-TMS-Fc 交联剂。研究还发现两个关键增效特征：二茂铁环取代基的相互作用，以及双环连接大体积基团时的高断裂响应性，用该交联剂制备的聚丙烯酸酯韧性显著提升，可延长材料寿命以减少废弃物^[15]。

1.5.6 小样本机器学习提升量子化学精度

密歇根大学开发出小样本机器学习提升量子化学精度的方法，核心在于逆向 DFT 技术的应用。研究团队从高精度量子多体理论数据中提取精确交换关联势，以此为训练信号培育出 NNLDAs 与 NNGGAs 两类神经网络泛函。这些模型无需直接拟合热化学数据，在小分子测试中平均绝对误差低于 2kcal/mol/ 原子，能改善电子密度的物理质量，为药物设计、电池材料研发等领域的量子模拟提供了精准工具^[16]。

1.5.7 E (3)等变基础模型实现十亿原子级分子动力学模拟

南加州大学的 Allegro-FM 模型凭借 E (3) 等变神经网络实现突破，可在超算上完成 40 亿原子系统的

分子动力学模拟，并行效率达97.5%，且覆盖种化学元素。该模型具备零样本学习能力，无需额外训练即可预测碳酸分解等复杂反应，过渡态能量误差仅0.3kcal/mol。其应用场景已延伸至碳封存混凝土开发、地幔矿物反应机制研究等领域，为宏观材料的微观模拟提供了全新范式^[14]。

1.5.8 DNA 神经网络实现分子尺度自主学习

该成果首次突破传统分子计算“无法学习”的局限，设计出含可激活放大门控、可激活转换门控的DNA神经网络，能像生物系统一样通过训练形成长期记忆。实验构建了包含700余种分子物种、1200余个独特链段的100位记忆网络，在72个测试中展现精准分类能力，且具备独立性、整合性、泛化性和稳定性四大关键特征。团队还通过“批训练”验证系统鲁棒性，不同训练顺序下结果一致，为人工分子智能走向实用化奠定基础，可用于精准医疗、智能材料等领域^[17]。

1.6 颠覆性意义

1.6.1 特征

与已有研究成果相比，AIChem平台是完全基于分子级及以下数据和化学变化数据的，而完全不依赖于人类告诉它已有的知识进行学习。依照于化学“结构决定性质，性质决定用途”的基本逻辑，并借助人工智能的可以推理的能力，在已知分子级及以下结构的条件下，AIChem平台可以完成对性质和反应的全推理过程。

AIChem平台与过去的平台相比，是智能化的算法逻辑。其根本存在方式并不在于“做实验”，而在于推理实验，即底层逻辑的彻底数据化。AI可以在没有思维惯性的情况下，打通思路，提供新的范式。

在应用端，AIChem平台可以在强大的算力和逻辑能力上甩开有着相似功能的人工智能，其准确性和完全不依赖经验的预言能力将使得所侦测的化学变化可以被提前计算得到，极大地加速相关产业变革。

1.6.2 理论科学方面

过去化学的研究方式是：人工做实验，不断试错，大浪淘沙^[3]。在AIChem平台的应用场景中，可以利用底层数据进行实验，颠覆了传统的人工实验或利用已知基本性质的AI做实验的方法。极大地节省了实验时间和耗材，大大提高了效率。

同时，由于AIChem平台是完全基于数据与运算逻辑的系统，因此，得到的结果由微观事实搭建而成，从而使得运算结果可以指示微观变化。如果将AIChem平台的运算结果与实际实验成果进行比较，便可以通过误差指示人类目前化学理论所构建的模型和真实微观粒子的差异，帮助人类重构理论体系，探秘微观奥秘。进而带动人类化学认知水平发展，从而更进一步地引发物理学、生物学、医学、

药学、材料学、化学工程、食品工程、能源动力工程、自动化工程、机械、仪器学等理工科甚至是法学、教育学等社会科学的大变革。

1.6.3 工程科学方面

在 AIChem 平台的应用过程中，首先受到影响的是材料科学与工程与制药工程。科技工作者可以通过 AIChem 平台搭建在目标条件下所需要的分子，从而由 AIChem 平台给出相应的经济合理型、环境温和型合成方式。或根据环境所需，由 AI 自主提出最为合适的材料或药品（可能为未曾提出的新型材料或药品），为人类的工程应用方面开拓视野。由此，可以为人类认知面的拓展做出贡献。这样建模得到的材料和药品更加的专门化、实用化，可以应对各种特殊或极端条件下的非典型问题或打通各种科技发展的痛点、堵点。

接下来受到影响的是化学工程与技术。在 AIChem 平台的广泛应用中，各种未有 AIChem 平台提出的方案优秀的制备方案将被慢慢取缔。人类将以更加环保、可持续的方案制备所需的化学药品。同时，随着新型材料、医药的发现，大批新型材料、新型医药的化工厂将拔地而起，一齐为化工的转型做出贡献。

下面影响的是机械、仪器学、自动化工程、医学、食品工程、能源动力工程等学科。随着大量新型医药、新型材料的使用，将迎来新型机械、仪器的精度与产能的大爆发。人类将可以利用新型科研仪器去做曾经无法做的实验，使得食品更加安全、医疗更加稳健、能源更加充足，人类将实现资源的充分利用。

最终，AIChem 平台的出现，将影响人类观念。法学、教育学将随着人类观念的进步而大步向前，对于政治、经济、文化产生深刻影响。

总之，AIChem 平台的应用，将带来一场以化学为开端的生产生活乃至社会历史的大变革。人类的生产力将得到极大的解放，生活极大得到富足，思想得到极大升华。

二、研究目标及阶段划分

2.1 研究总目标

在对已有研究的基础上，利用人工智能，搭建基于基本微观化学数据和化学物质组合规则的算法平台（即基于理化数据推演的人工智能化学平台），实现人工智能通过相关数据推算物质结构和理化性质，进一步推算化学反应。从而在化学相关产业上实现从研发到生产的全流程 AI 赋能，推动人类理论化学和以化学为基础的后端工程科学的发展。

2.2 研究阶段

第一阶段：确立 AIChem 平台研究项目，确定 AIChem 平台功能体系，初步了解市场需求，完成

初步研发端全流程设计。

第二阶段：研究已有资料，进一步深化 AIChem 平台的特色，使之能够清晰地有别于已有的人工智能化学建模系统，填补技术空白。

第三阶段（研发 I 阶段）：开始研发并导入原子级及以下微观层级的数据，训练人工智能使之对已由或未发现的物质存在性、理化性质及宏观表现的判断具有较高准确率。

第四阶段（研发 II 阶段）：导入化学反应数据，训练人工智能使之对已由或未发现的化学反应的存在性、反应现象宏观表现和各产物产率的判断具有较高准确率。

第五阶段（应用 I 阶段）：对项目成果进行宣发、演示，尝试投入不同应用场景，进一步开发潜在价值；

第六阶段（应用 II 阶段）：对项目应用成果进行专门化调试，以适配应用场景下的各项成本需求，以期带来更大价值；

第七阶段：投入实际市场应用。

三、挑战性与不确定性

3.1 挑战性

3.1.1 算法与模型

在技术层面上，人工智能化学平台想要达到应用层面，必须具有较强的模型可行性。而现有 AI 智能模型翻译比较表面化，难以涉及底层逻辑，且推导过程中前后连贯性较差，不能提供实质性建议。同时，在 AI 智能模型翻译上，有技术、科学、逻辑等多重挑战。如何让 AI 真正“理解”和“创造”化学组合是核心挑战^[18]。这方面的问题包括但不限于：

（1）解决分子与组合的表示问题

将一个分子（甚至是未知的分子）有效地转化为计算机可以理解的表示（即特征工程）是 AIChem 最核心的研究内容。其关键挑战在于“组合”的表示。“组合”不是单个分子的简单加和，而是建模分子间的相互作用（如氢键、共轭效应与超共轭效应、静电作用等）的叠加^[18]。

（2）突破高维但已知知识密度稀疏的化学空间不易

可能的化合物组合数量多，所形成的空间维度高。人类目前探索过的只是其中极小、极稀疏的一点。AI 模型很容易在这种空间中外推失败，产生不准确的预测^[18]。

（3）模型的可释性转化存在困难

化学家需要相信模型的预测。如果一个 AI 模型只是给出一个“黑箱”预测（如“组合 A 和 B 有 80% 的概率形成稳定共晶”）而无法解释原因，其建议是无法被采纳的。发展具有可解释性的 AI (XAI)

对于化学领域的应用至关重要^[18]。

Optimization Method	Optimization Target(s)	Learning Setting(s)	Search Scope
Stochastic Gradient Descent (SGD)	Model parameter optimization	Supervised learning	Local optimization
Adam	Model parameter optimization	Supervised learning	Local optimization
Backpropagation	Gradient computation (for model training)	Supervised learning	Supports local optimization (via optimizers)
Monte Carlo (MMC, SLHMC)	Molecular optimization	Reinforcement learning, Unsupervised	Global optimization
Bayesian Optimization	Hyperparameter, Molecular optimization	Supervised, Reinforcement	Global optimization
Graph-based Models	Molecular optimisation	Supervised, Unsupervised	Local (parameter tuning) and global (topology-driven)

图4：化学机器学习中的优化策略及其应用总结^[18]

(4) 达到多目标优化成果复杂

在探索组合时，目标往往不是单一的(如我们可能同时希望一个催化剂组合具有高活性、高选择性和长寿命)。这些目标通常是相互冲突的。AI平台需要能够进行多目标优化，提供一系列帕累托最优解，而不是单个答案^[18]。

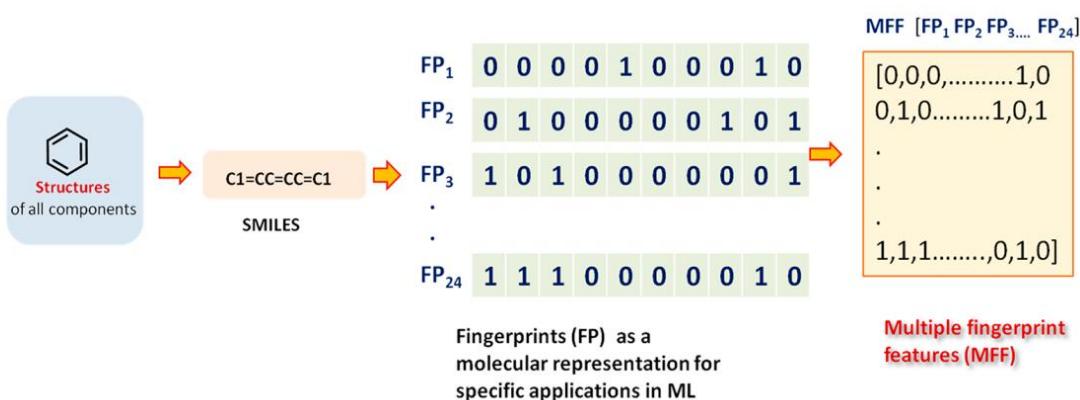


图5：苯环数字化举例^[19]

(5) 主动学习与贝叶斯优化过程复杂

由于实验所需的时间成本、人力成本较高，AIChem 平台必须能够智能地决定“下一个实验做什么”，以用最少的实验次数获得最大的信息增益。这通常依赖于主动学习和贝叶斯优化，但其效

率高度依赖于不确定性估计的准确性^[18~19]。

3.1.2 海量数据集成

人工智能化学平台的训练需要海量的数据，而在其海量的数据分层中，很容易出现数据误差；同时，在数据中关系链交错复杂的情况下，在海量数据中容易出现偶然性的逻辑谬误。因此在技术方面对算法要求较高，具有很强挑战性。这些挑战包括但不限于：

(1) 数据稀缺与高成本

高质量的化学实验数据，尤其是关于“组合”效应的数据，在已知的化学体系中占比较少，且获取成本极高。进行一次合成、分离、表征可能需要数小时甚至数天，而高通量实验设备价格昂贵，是获取数据的最大障碍。

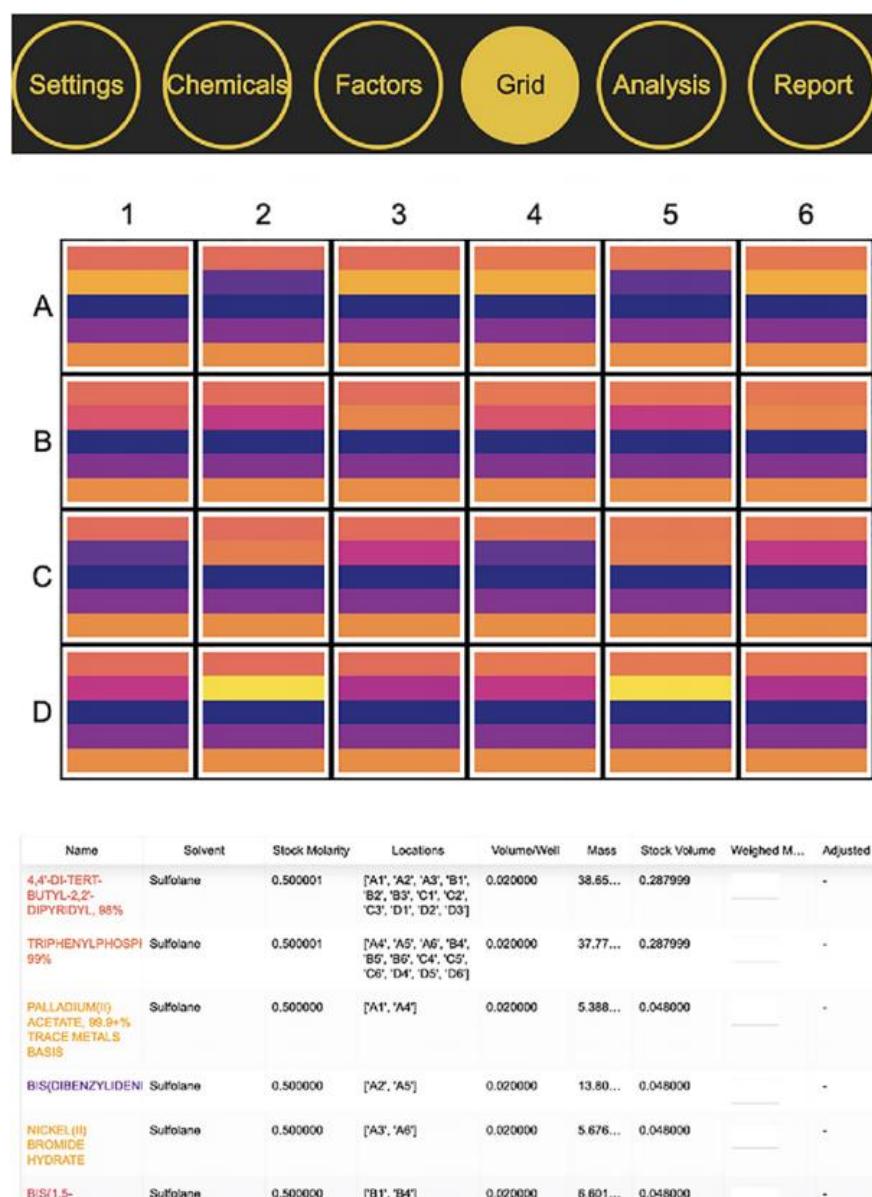


图 6：利用 phactor 进行高通量试验设计^[20]

(2) 数据不一致与标准化

数据来自不同实验室、不同仪器、不同操作人员，鲜有统一标准。在表征数据的处理流程存在差异，导致不同研究标准的数据进入使得结果更加不精确^[20]。

(3) 数据的多维性与复杂性

化学数据不仅是数字表格。它包括但不限于：

结构化数据：产率、温度、浓度等；

非结构化数据：实验记录、论文文本；

谱图数据：需要复杂的解析的 NMR、IR、质谱；

时空数据：反应过程的实时监测数据。

综合考虑这些数据需要强力的算法和算力，以及人工智能的强推理能力。在这样的多维数据的情况下，结果才能愈加精确^[20]。

(4) “负结果”报告缺失

失败的实验（即“组合”不成立的情况）通常不会被发表或记录，但这对于 AI 模型理解化学空间的边界至关重要。缺乏负结果会导致模型产生严重的偏见和过度的乐观预测^[21]。

3.1.3 实验自动化与闭环挑战

目前，最为理想的 AI 化学平台是一个“AI 驱动-机器人执行-自动反馈”的闭环系统。但是，由于以下原因，任何一个环节的故障都可能导致整个实验链中断，甚至造成研发事故。

1. 自动化实验的复杂性（只有部分化学实验易于自动化甚至智能化）^[21]；

2. 高通量表征的瓶颈（可以高通量合成，但表征的速度较慢，可能成为流程中的瓶颈）^[21]；

3. 闭环的可靠性存疑（实现全自动化的“设计—合成—测试—分析”循环，需要极高的硬件可靠性和软件稳定性）^[21]；

4. 软硬件接口与集成问题（将来自不同供应商的自动化设备、分析仪器和 AI 软件平台无缝集成是一个巨大的工程挑战，需要统一的通信协议和数据标准）^[21]。

相比于目前的 AI 化学实验机器人，AIChem 平台对真实实验场景要求较低，但是对于数据获取的要求和实验准确性的理解更高。这对 AIChem 平台智能化研发提出极高要求，实现数字化实验闭环就更富有挑战性。

3.1.4 领域知识整合挑战

AI 不能脱离化学的物理定律和专家知识，在设计平台中需要专业融合，以及专家配合来纠正模型错误，促进平台向正确推理方向发展。这一纠正过程中，也需要已有实验数据进行比对，需要多领域

交叉融合。

3.1.5 强大的基础设施与硬件条件

化学平台在运行中会产生海量数据，需要大量的GPU计算资源以及一个强大的，可扩充的且安全的数据管理系统。而这些意味着强大的基础设施与硬件条件，对于目前世界任何一个国家，这是一项富有挑战性的工作。

3.2 不确定性

3.2.1 未来技术发展的不确定

AI模型创造力的上限未知，它的可应用性到哪个程度至今还是谜题。

目前的弱人工智能算力不支持AIChem平台的算力要求，若AGI能够实现，可以大大加快AIChem研发进程；其次，在化学组合过程中，复杂系统使性质的可预测性存疑；另外，未来算法范式可能有颠覆性变革，现有平台算法结构可能有突破性架构。

3.2.2 科学与哲学的不确定

化学平台在决策过程中可能出现黑箱问题，这带来哲学上的不确定性。化学未来发展进程是否与AIChem平台同轨存疑。此外，它还有可能对化学空间发展产生影响，从而被限制^[22]。

3.2.3 经济系统不确定性

由于投资回报率的不确定性，可能导致资本与社会对平台的支持减弱，从而影响平台的产生。此外，知识产权归属的模糊性也会影响平台的发展^[23]。

3.2.4 人才协作的不确定

未来化学家的角色定位很有可能发生改变，与之相对的平台功能也会存在问题，进而影响到平台发展方向。

3.3 现状综合分析和问题解决方案总思路

3.3.1 现状综合分析

首先，在源头探索与设计层面，研究者在进行研究的过程中，如同“大海捞针”。化学空间理论上近乎无限，如何从数以亿计的候选分子中精准定位兼具优异性能和可制备性的目标，是首要难题。

其次，在实验验证与制备环节，挑战从虚拟走向现实。即便理论预测成功，在实验室中将其合成出来也极具风险。所需的反应条件可能极为苛刻（如高温高压），前驱体原料昂贵或稀有，且合成步骤冗长、产率极低。更棘手的是，许多新物质（如新材料或催化剂）在宏观尺度的表现与理论计算或微观测试结果迥异，可能出现性能衰减或不稳定，难以达到预期效果。

最后，在走向应用与产业化阶段，经济性与安全性成为最大壁垒。一个能在实验室克级制备的

样品，与工业化吨级生产之间存在巨大鸿沟。放大效应会引入新的复杂性，如传质、传热不均，导致产品质量不稳定。同时，全生命周期的评估至关重要，包括其毒性、环境影响以及生产过程中的安全性，任何一项不达标都会令之前的所有努力前功尽弃。

3.3.2 问题解决方案总思路

其一，运用大数据与AI建立数据库，导入已知的化学组分，形成“源”，初步利用智能进行简易组合，输入下一个库中进行筛选，提取可存在的，性质良好的化合物。

其二，AI通过训练海量化学数据，能高精度预测化合物性质。输入分子结构，AI模型可快速输出其毒性、溶解度、稳定性及反应活性等关键参数，大幅节省实验成本与时间。并且能够模拟其在高温高压等现实实验室难以达到的特殊环境进行实验，还可统计规律，预测其在长时间后是否发生性质变化。

其三，AI通过分析海量数据，能前瞻性评估化工原料的毒性、反应风险及存储稳定性。它可预测不同条件（如加热、混合）下的潜在危险，智能推荐更安全的替代品或优化工艺条件，从源头提升本质安全水平。同时，AI驱动的实时监控能及时预警异常，有效预防事故，为化工安全生产提供关键决策支持。

四、主要研究内容及初步研究方案

4.1 主要研究内容

在算法逻辑设计方面，主要研究包括但不限于：数据库建立及其内部模块设置、逻辑设置、自我纠错算法、检索库设置；

在数据库数据导入方面，主要研究包括但不限于：导入数据来源、数据清洗与标准化方法；

在硬件设施方面，主要研究包括但不限于：关键计算网络、关键技术手段和测量手段。

4.2 总体研究框架

该“基于理化数据推演的人工智能化学平台”流程分为研发阶段与应用阶段，具体如下：

研发阶段：数据驱动的化学体系构建

基础数据库建设：构建「原子级以下数据库」「原子级数据库」，为「化学物质库」提供底层理化数据支撑；同时，基础数据库与化学物质库之间通过纠错机制持续校验、优化数据准确性。

化学反应库生成与迭代：以「化学物质库」为基础，依托「化学反应数据库」，通过机器学习生成「化学反应库」；且「化学物质库」与「化学反应库」之间、「化学反应库」与「化学反应数据库」之间，均通过纠错循环不断完善化学反应相关数据（如物质反应活性、反应条件等）。

产品研发输出：经过多轮数据迭代与优化，从「化学反应库」逐步导向产品的研发与产出。

应用阶段：多领域技术转化

研发阶段构建的人工智能化学平台，将成果应用于化工、药学、材料学、医学等多个学科领域，实现技术的落地转化与产业价值释放。

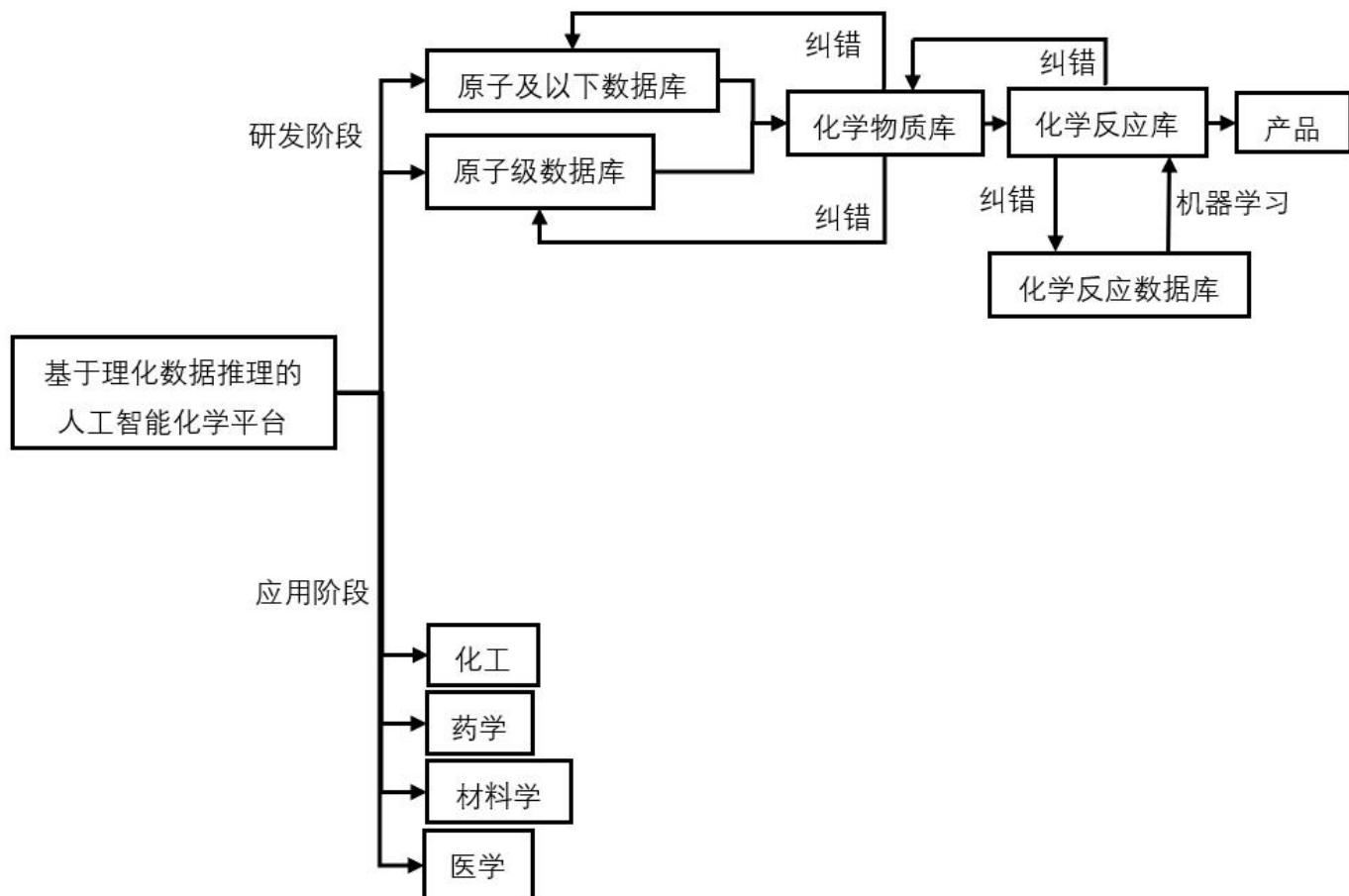


图 7：总体框架设计

4.3 算法逻辑分析

4.3.1 整体算法设置

整体算法采取“简流程、多纠错、可兼容”的基本逻辑，对算法逻辑进行设计：

简流程：是指在保证系统基本性能的基础上，尽可能减少数据库的堆砌，使算法整体趋向于简单，以免浪费算力；

多纠错：是指在保证系统基本性能和流程简便的基础上，为保证系统准确性，应多次对平台结论进行判断，使之符合客观事实；

可兼容：是指在保证系统基本性能、流程简便和正确性的基础上，整合化学学术体系，使得有机化学、无机化学、分析化学等相关化学理论能够在同一系统中体现，使该系统成为完全的“化学研究百科全书”。

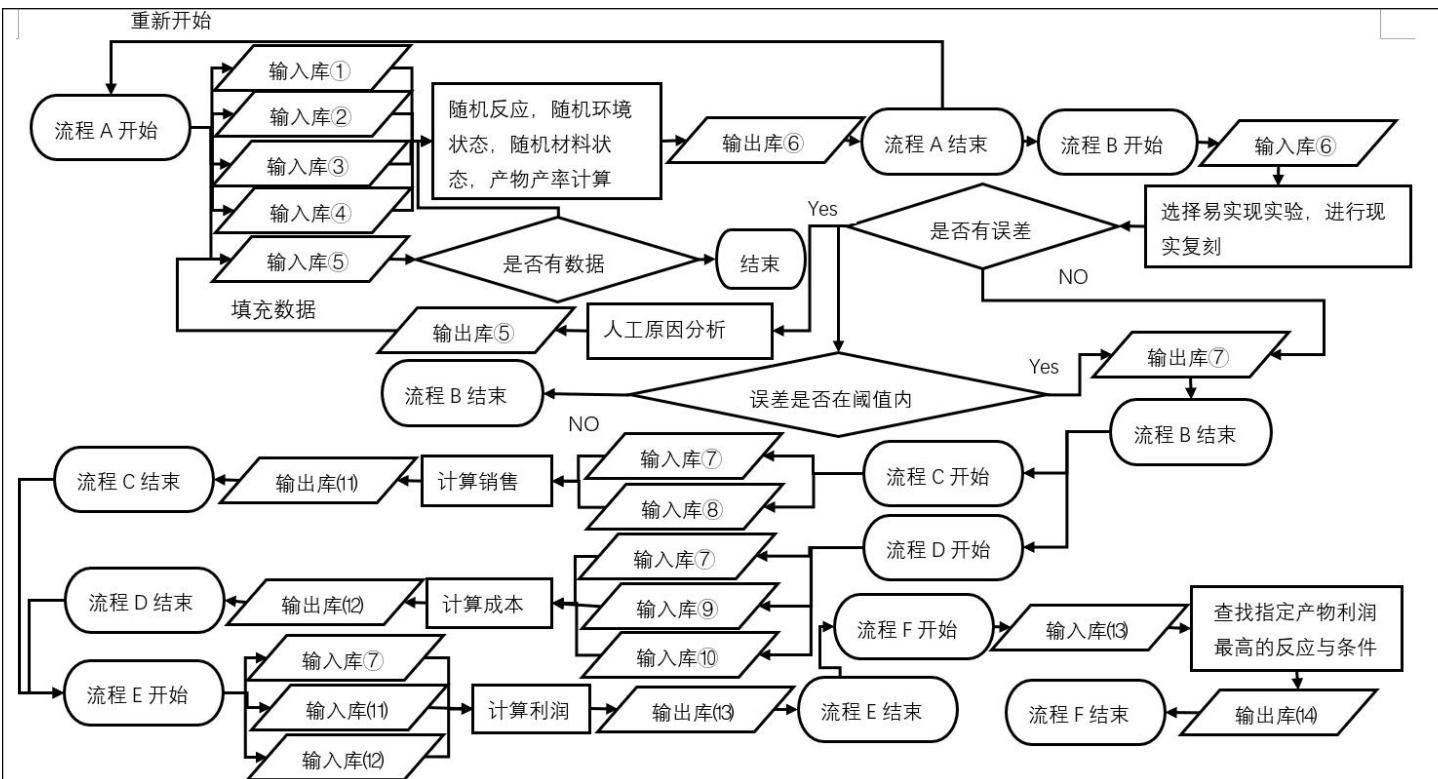


图 8：算法整体流程设计

(1) 无机化合物训练流程

图例和注记

库①：原子数据库含有半径，电子数等数据

库②：原子间影响库含有化学键长，角度等因素

库③：影响补充库含有库②中未考虑到的影响因素

库④：分子数据库含有随机组合成的分子的名称，角度，键长等数据

库⑤：正确分子数据库含有现实中已确认的分子名称，角度，键长等数据

库⑥：名称一致库含有库④中与库⑤名称一致的分子的数据

库⑦：名称不一致库，含有库④中与库⑤名称不一致的分子的数据

库⑧：矛盾库含有库⑥中与库⑤数据矛盾的分子的数据

库⑨：不矛盾库含有库⑥中与库⑤数据不矛盾的分子的数据

库⑩：合成方法库，含有现实中已确认的分子合成方法数据

库⑪：合成方法补充库含有新分子的合成方法数据

库⑫：可尝试应用库，还有可能成立的新分子数据及合成方法

流程叙述

流程 A：原子组合生成分子

输入与数据填充：利用「库①（原子数据库，含半径、电子数等）」「库②（原子间影响库，含键长、角度等）」「库③（影响补充库，补充库②未考虑的因素）」填充数据。

原子随机组合：若输入库有数据，原子按规则随机组合，生成的分子数据（名称、键长、角度等）存入「库④（分子数据库）」；若输入库无数据，流程直接结束。

数据质量校验：填充数据后，校验“是否成功”“错误率是否 $<2\%$ ”。若错误率过高，删除数据；若需人工分析原因，结果存入「库⑧（矛盾库）」或「库⑨（不矛盾库）」。

流程 B：分子与已知正确分子的一致性验证

名称比对：流程 A 结束后，将「库④（生成的分子）」与「库⑤（正确分子数据库，含现实确认的分子数据）」比对分子名称：

名称一致：数据存入「库⑥（名称一致库）」；

名称不一致：数据存入「库⑦（名称不一致库）」。

数据量与矛盾校验：

对「库⑥」，校验“数据量是否达 200”；同时比对「库⑥」与「库⑤」的数据（键长、角度等）是否矛盾：矛盾存入「库⑧」，不矛盾存入「库⑨」。

若“名称不一致”（库⑦）且错误率 $<2\%$ ，则进入后续「流程 C」；否则删除数据，流程 B 结束。

流程 C：新分子的合成方法推理

针对「库⑦（名称不一致的潜在新分子）」，结合「库⑩（合成方法库，含现实确认的合成方法）」，尝试推理分子合成方法：

自动推理：若推理成功，新分子数据及合成方法存入「库⑫（可尝试应用库）」，合成方法补充存入「库⑪（合成方法补充库）」，流程 C 结束。

人工推理：若自动推理失败，进入人工推理；若人工推理仍失败，且“推理次数达 20 次”，则删除数据，流程 C 结束；若人工推理成功，结果同样存入「库⑫」和「库⑪」。

核心逻辑

从原子级数据出发，通过“随机组合生成分子→与已知正确分子比对→对潜在新分子推理合成方法”的流程，逐步筛选、探索“可能成立的新分子及其合成方法”，各“库”作为数据载体，支撑不同阶段的判断与迭代。

（2）有机化合物原子到残基流程分析

图例和注记

库①：原子数据库含有半径，电子数等数据

- 库②：原子间影响库含有化学键长，角度等因素
- 库③：影响补充库含有库②中未考虑到的影响因素
- 库④：残基数据库含有随机组合成的残基的名称，角度，键长等数据
- 库⑤：正确残基数据库含有现实中已确认的残基名称，角度，键长等数据
- 库⑥：名称一致库含有库④中与库⑤名称一致的残基的数据
- 库⑦：名称不一致库，含有库④中与库⑤名称不一致的残基的数据
- 库⑧：矛盾库含有库⑥中与库⑤数据矛盾的残基的数据
- 库⑨：不矛盾库含有库⑥中与库⑤数据不矛盾的残基的数据
- 库⑩：合成方法库，含有现实中已确认的残基合成方法数据
- 库⑪：合成方法补充库含有新残基的合成方法数据
- 库⑫：可尝试应用库，还有可能成立的新残基数据及合成方法

流程叙述

流程 A：原子组合生成残基

数据输入与填充：利用「库①（原子数据库，含原子半径、电子数等）」「库②（原子间影响库，含化学键长、角度等）」「库③（影响补充库，补充库②未覆盖的影响因素）」填充数据。

原子随机组合：若输入库有数据，原子按规则随机组合，生成的残基数据（名称、键长、角度等）存入「库④（残基数据库）」；若输入库无数据，流程直接结束。

数据质量校验：填充数据后，校验“是否成功填充”“错误率是否 $<2\%$ ”：

若需人工分析原因，结果存入「库⑧（矛盾库，残基数据与正确残基矛盾）」或「库⑨（不矛盾库，残基数据与正确残基一致）」；

若错误率 $\geq2\%$ ，删除数据，流程 B 结束；若错误率 $<2\%$ ，进入流程 B。

流程 B：残基与已知正确残基的一致性验证

名称比对：流程 A 结束后，将「库④（生成的残基）」与「库⑤（正确残基数据库，含现实确认的残基数据）」比对**残基名称**：

名称一致：数据存入「库⑥（名称一致库）」；

名称不一致：数据存入「库⑦（名称不一致库）」。

数据量与矛盾校验：

对「库⑥」，校验“数据量是否达 200”；同时比对「库⑥」与「库⑤」的数据（键长、角度等）是否矛盾：矛盾存入「库⑧」，不矛盾存入「库⑨」。

若“名称不一致”（库⑦）且错误率 $<2\%$ ，则进入后续「流程 C」；否则删除数据，流程 B 结束。

流程 C：新残基的合成方法推理

针对「库⑦（名称不一致的潜在新残基）」，结合「库⑩（合成方法库，含现实确认的残基合成方法）」，尝试推理残基合成方法：

自动推理：若推理成功，新残基数据及合成方法存入「库⑫（可尝试应用库）」，合成方法补充存入「库⑪（合成方法补充库）」，流程 C 结束。

人工推理：若自动推理失败，进入人工推理；若人工推理仍失败，且“推理次数达 20 次”，则删除数据，流程 C 结束；若人工推理成功，结果同样存入「库⑫」和「库⑪」。

核心逻辑

从原子级数据出发，通过“随机组合生成残基→与已知正确残基比对→对潜在新残基推理合成方法”的流程，逐步筛选、探索“可能成立的新残基及其合成方法”，各“库”作为数据载体，支撑不同阶段的判断与迭代。

（3）有机化合物残基到分子训练流程

图例和注记

库①：残基数据库含有键长，角度等数据

库②：残基间影响库含有特殊化学键等因素

库③：影响补充库含有库②中未考虑到的影响因素

库④：分子数据库含有随机组合成的分子的名称，角度，键长等数据

库⑤：正确分子数据库含有现实中已确认的分子名称，角度，键长等数据

库⑥：名称一致库含有库④中与库⑤名称一致的分子的数据

库⑦：名称不一致库，含有库④中与库⑤名称不一致的分子的数据

库⑧：矛盾库含有库⑥中与库⑤数据矛盾的分子的数据

库⑨：不矛盾库含有库⑥中与库⑤数据不矛盾的分子的数据

库⑩：合成方法库，含有现实中已确认的分子合成方法数据

库⑪：合成方法补充库含有新分子的合成方法数据

库⑫：可尝试应用库，还有可能成立的新分子数据及合成方法

流程叙述

流程 A：残基组合生成分子

数据输入与填充：利用「库①（残基数据库，含残基键长、角度等数据）」「库②（残基间影响

库，含特殊化学键等因素）」「库③（影响补充库，补充库②未覆盖的影响因素）」填充数据。

残基随机组合：若输入库有数据，残基按规则随机组合，生成的分子数据（名称、键长、角度等）存入「库④（分子数据库）」；若输入库无数据，流程直接结束。

数据质量校验：填充数据后，校验“是否成功填充”“错误率是否 $<2\%$ ”：若需人工分析原因，结果存入「库⑧（矛盾库，分子数据与正确分子矛盾）」或「库⑨（不矛盾库，分子数据与正确分子一致）」；若错误率 $\geq 2\%$ ，删除数据，流程 B 结束；若错误率 $<2\%$ ，进入流程 B。

流程 B：分子与已知正确分子的一致性验证

名称比对：流程 A 结束后，将「库④（生成的分子）」与「库⑤（正确分子数据库，含现实确认的分子数据）」比对分子名称：

名称一致：数据存入「库⑥（名称一致库）」；

名称不一致：数据存入「库⑦（名称不一致库）」。

数据量与矛盾校验：

对「库⑥」，校验“数据量是否达 200”；同时比对「库⑥」与「库⑤」的数据（键长、角度等）是否矛盾：矛盾存入「库⑧」，不矛盾存入「库⑨」。

若“名称不一致”（库⑦）且错误率 $<2\%$ ，则进入后续「流程 C」；否则删除数据，流程 B 结束。

流程 C：新分子的合成方法推理

针对「库⑦（名称不一致的潜在新分子）」，结合「库⑩（合成方法库，含现实确认的分子合成方法）」，尝试推理分子合成方法：

自动推理：若推理成功，新分子数据及合成方法存入「库⑫（可尝试应用库）」，合成方法补充存入「库⑪（合成方法补充库）」，流程 C 结束。

人工推理：若自动推理失败，进入人工推理；若人工推理仍失败，且“推理次数达 20 次”，则删除数据，流程 C 结束；若人工推理成功，结果同样存入「库⑫」和「库⑪」。

核心逻辑

从获得的残基级数据出发，通过“残基随机组合生成分子→与已知正确分子比对→对潜在新分子推理合成方法”的流程，逐步筛选、探索“可能成立的新分子及其合成方法”，各“库”作为数据载体，支撑不同阶段的判断与迭代。

4.4 研发 I 阶段数据导入

分析化学与计算化学是化学学科的两大重要分支，前者是“化学的眼睛”，聚焦通过实验检测解析物质组成与结构；后者是“化学的大脑”，依托理论与计算机模拟探索化学本质，二者协同推动化

学研究从实验驱动向“实验+理论”双轮驱动发展。在研发Ⅰ阶段，最核心的是两个数据库，进行物质的分析。

物质原理数据库

4.4.1 元素周期律库

数据包括但不限于：元素基本信息（原子序数、符号、原子量）、电子构型（基态电子排布式、价电子构型）、周期表定位（周期/族/区）、周期性参数（电离能、电子亲和能、电负性、原子半径、离子半径等数据）。

对于新物质与原有数据符合度较高的，类比元素周期律给出提示，之后进入AIChem平台模拟实验。对于关联度低的新物质，即与之前发现的物质在理化性质方面的组合不同，该系统会自动收集相关理化性质，并形成性质间的关联，之后进入模拟实验系统。

表1：部分元素数据列举^[24~26]

元素	原子序数	族	核外电子构型	第一电离能/(kJ • mol-1)	第一电子亲和能/(kJ • mol-1)	电负性(鲍林标度)
H	1	IA族	1s ¹	1312.0	-73.0	2.20
C	6	IVA族	[He]2s ² 2p ²	1086.5	-122.0	2.55
O	8	VIA族	[He]2s ² 2p ⁴	1313.9	-141.0	3.44
Na	11	IA族	[Ne]3s ¹	495.8	-53.0	0.93
Mn	25	VIIIB族	[Ar]3d ⁵ 4s ²	717.3	-50.0	1.55
Fe	26	VIII族	[Ar]3d ⁶ 4s ²	759.4	-44.0	1.83
Cu	29	IB族	[Ar]3d ¹⁰ 4s ¹	745.5	-119.0	1.90
Xe	54	0族	[Kr]4d ¹⁰ 5s ² 5p ⁶	1170.4	+40.0	2.60
Ra	88	IIA族	[Rn]7s ²	509.3	+10.0	0.97

4.4.2 化学键理论库

相关数据包括但不限于：化学键类型（离子键、共价键、金属键、氢键、范德华力等）、键参数（键长、键能、键角，标注测定条件如常温常压）、成键规律（如离子键形成条件-电负性差>1.7，共价键饱和性与方向性）、特殊成键案例（金属-金属多重键、大π键等非经典化学键的成键原理与特征）。

表2：部分物质化学键数据列举^[27~31]

化学键名称及类型	离子性百分数 /%	键长/pm	键能/(kJ •mol-1)	键角/°
O-H 极性共价键	32.0	96	464	104.5
Na ⁺ 与 Cl ⁻ 的离子键	71.2	282	786	无
S-H 极性共价键	3.5	134	347	92.1
Al-Cl 极性共价键	45.1	221	318	101、110
K ⁺ 与 Cl ⁻ 离子键	54.1	312	1800	无
Ga-N 极性共价键	31.5	195	370	109.5
C-C 非极性共价键	0	142	413	120

4.5 研发II阶段数据导入

在经过 AIChem 平台的计算后，物质的氧化性、还原性、酸碱性以及相关化学性质可以得到解决。然而，分子间作用力和相关人为定义的公式尚未完全解决。因此，需要在 II 阶段研发过程中加入可以表征上述两方面的子库，即物质状态与相变理论子库和理化公式子库。

4.5.1 物质状态与相变理论子库

相关数据包括但不限于：物质三态（气态、液态、固态）的判定标准（温度/压力范围）、相变类型（熔化、汽化、升华、晶型转变等）、相变热力学参数（熔点、沸点、相变焓、相变熵，标注标准大气压下数据）、极端条件相变规律（高温/高压/强磁场下的相变临界条件与理论解释）。

在考虑过程中，应注意气压对不同温度下物质的熔沸点的影响。

另外，由于不同温度将影响较大，还应当考虑不同物质在三相点、超临界点等特殊相变点的数据。

表3：部分物质化学键数据列举^[32~37]

物质名称	熔点/°C	沸点/°C	三相点/°C	超临界点（临界温度）/°C
水 (H ₂ O)	0.0 (1atm)	-100.0 (1atm)	0.01 (固液气平衡温度，对应压力 611.657Pa)	373.946 (超临界状态临界温度)

氯化钠 (NaCl)	800.7 (1atm)	1465 (1atm)	800.7 (接近熔点, 对应压力约 30Pa, 离子晶体三相点压力较高)	2800 (文献估算值, 高温下离子晶体临界温度)
乙醇 (C ₂ H ₅ OH)	-114.1 (1atm)	78.37 (1atm)	-114.1 (对应压力约 1.3×10^{-3} atm)	240.7 (超临界乙醇临界温度)
二氧化碳 (CO ₂)	-56.6 (5.11atm, 1atm 下为升华, 无液态沸点)	-78.5 (1atm, 升华点, 非液态沸点)	-56.6 (固液气平衡温度, 对应压力 5.11atm)	31.04 (超临界 CO ₂ 常用临界温度)
甲烷 (CH ₄)	-182.5 (1atm)	-161.5 (1atm)	-182.5 (对应压力约 0.11atm)	-82.6 (超临界甲烷临界温度)
氧气 (O ₂)	-218.8 (1atm)	-183.0 (1atm)	-218.8 (对应压力约 1.5×10^{-3} atm)	-118.6 (超临界氧气临界温度)
氮气 (N ₂)	210.0 (1atm)	-195.8 (1atm)	-210.0 (对应压力约 0.12atm)	-146.9 (超临界氮气临界温度)
蔗糖 (C ₁₂ H ₂₂ O ₁₁)	186 (加热至熔点后分解, 无稳定液态)	无 (未达沸点即分解)	无 (分解导致无明确固液气三相平衡)	无 (未达超临界状态即分解)
铝 (Al)	660.4 (1atm)	2467 (1atm)	660.4 (对应压力约 1atm, 金属三相点压力接近标准大气压)	7591 (文献估算的金属铝临界温度)
汞 (Hg)	-38.83 (1atm)	356.7 (1atm)	-38.83 (对应压力约 1.7×10^{-4} atm)	1477 (超临界汞临界温度)
氨 (NH ₃)	-77.73 (1atm)	-33.34 (1atm)	-77.73 (对应压力约 0.06atm)	132.4 (超临界氨临界温度)

以上步骤进行结束后, AIChem 平台将对相关物质的存在性和性质进行推算, AIVPBD 会比对原有物质, 对相关物质进行模拟实验, 最后, 将实验数据输入到下一个数据库。

4.5.2 理化公式子库

理化公式子库包括但不限于有关物理化学、分析化学、有机化学的重要公式，代表如下：

表4：部分理化公式

公式名称	公式内容
盖斯定律	$\Delta_rH_m = \sum \Delta_rH_{m,products} - \sum \Delta_rH_{m,reactants}$
吉布斯—霍姆赫兹方程	$\Delta_rG_m = \Delta_rH_m - T\Delta_rS_m$
标准摩尔反应焓	$\Delta_rH_m = \sum v_i \Delta_rH_{m,i,products} - \sum v_j \Delta_rH_{m,j,reactants}$
平衡常数	$K = [A]^a[B]^b/[C]^c[D]^d$
能斯特方程	$\psi = \psi^\ominus + (2.303RT \lg([Ox]/[Red]))/nF$
电极电势间接计算	$-n_3 F \psi^\ominus_3 = -n_1 F \psi^\ominus_1 - n_2 F \psi^\ominus_2$
范德霍夫等温式	$\Delta_rG_m = \Delta_rG_m^\ominus + RT \ln Q \quad (Q = [A]^a[B]^b/[C]^c[D]^d)$
理想气体状态方程	$pV = nRT$ 或 $pM = \rho RT$ 或 $pV_m = RT$
实际气体状态方程	$(p + a/V_m^2)(V_m - b) = RT$

表5：实际气体状态方程范德华常数列举^[38~41]

气体名称	范德华常数 a/(Pa·m ⁶ ·mol ⁻²)	范德华常数 b/(m ³ ·mol ⁻¹)
氢气	0.02476	0.02661×10^{-3}
氧气	0.1378	0.03183×10^{-3}
氮气	0.1408	0.03913×10^{-3}
二氧化碳	3.640	0.04267×10^{-3}
氨气	4.225	0.03707×10^{-3}
甲烷	0.2283	0.04278×10^{-3}
水蒸气 (273.15K)	5.536	0.03049×10^{-3}
氯气	6.579	0.05622×10^{-3}

在由理化公式提供吉布斯自由能，电极电势，焓变等相关理化计算公式后，AIChem 平台可以分析该物质可能参与的化学反应，并得出该反应的全部化学信息。

由此，AIChem 平台的研发工作到此结束。此刻已经为 AIChem 平台的应用做出了一切准备。

4.6 硬件设施

4.6.1 技术手段（4 个三级节点）

光谱/能谱分析：原子发射光谱（定元素组成）、XPS（定价态）、拉曼光谱（定化学键类型）、红外光谱（确定电子结构）；

结构成像：冷冻 TEM（原子排列观测）、XRD（晶体结构和物相分析）；

电化学初筛：四探针法（测导电性）、循环伏安法（测氧化还原活性）；

磁学初筛：振动样品磁强计（VSM，测磁学性质，反推电子自旋状态）。

4.6.2 核心目标

锁定原子间关系（化学键类型、键长和键能、键级、原子排列规律）。

五、预期成果及转移转化设想

5.1 预期成果

AIChem 平台是一个以人工智能为核心、以理化数据为基础、以推演与优化为手段的综合性化学研究与应用系统。它不仅是化学研究的“智能大脑”，也是连接理论研究、实验验证与产业应用的桥梁。其具有分子结构与性质预测，化学反应路径推演，新物质发现与合成方法设计，物质多目标优化与实验设计，与其它试验系统对接等功能。

其在科学方面的预期成果有：

- (1) 推动化学理论发展：通过对比模拟与实验结果，验证或修正现有化学理论；
- (2) 发现新物质与新反应：系统性地探索未知化学空间，发现具有特殊性质的新分子、新材料；
- (3) 加速药物与材料研发：在医药、能源、材料等领域实现高效、精准的分子设计与筛选；
- (4) 降低研发成本与周期：大幅减少实验试错次数，提升研发效率。

其在工程应用方面的预期成果有：

- (1) 优化合成路径，降低能耗与污染；
- (2) 实现从实验室到产业化的高效转化，提升化工生产的智能化水平；
- (3) 结合 AI 进行药物成分分析、药物设计与个性化医疗方案推演。

5.2 成果应用可行性和预期应用场景

5.2.1 经济可行性

建设一个完整的 AI 化学平台需要巨额前期投资，包括：

- (1) 研究硬件（自动化机器人臂、液体处理器、反应器阵列、高通量表征设备），单件设备可能高达数十万至数百万美元，价格高昂。
- (2) 软件：AI 模型开发与训练、数据管理平台、控制系统集成，需要顶尖的 AI 专家和软件工

程师团队。

(3) 基础设施（专用的实验室空间和配套设施）与运营成本（维护费用、化学品消耗、计算资源（云服务或服务器集群）、专业团队的薪资）

其潜在的经济回报也十分丰厚，有以下方面：

节省巨大的时间成本，如在制药行业，药物晚上市一天的损失可能高达百万美元。平台能将数年的人工实验缩短至数月

降低实验成本，通过 AI 的精准预测和主动学习，能将需要进行的实验数量降低 1-2 个数量级，直接节省了人力、物料和时间成本。

发现高价值产品，平台有能力发现人类研究者容易忽略的“非直觉”材料或分子，从而催生“重磅炸弹”药物、高性能电池材料、革命性催化剂等，带来垄断性利润。

优化现有工艺，降低成本。如在化工行业，对现有生产工艺（如催化剂配方、反应条件）进行微小优化，就能带来显著的能耗降低和产率提升，产生持续的现金流。

由此得出结论：

目前对大型企业和顶级研究机构可行：对于财力雄厚的跨国药企（如辉瑞、罗氏）、大型化工企业（如巴斯夫、陶氏）和国家实验室，这被视为一项维持未来竞争力的战略性投资。他们有能力承担初期的亏损，以换取长期的技术壁垒。

对中小企业是挑战：对于初创公司和中小型企业，自建完整平台的门槛过高。但这催生了新的商业模式：

平台即服务：专门的公司（如 Insilico Medicine、Atomwise 在药物发现领域）提供基于其 AI 平台的计算和实验服务，客户按项目付费。

在未来，随着硬件成本下降和开源软件生态的成熟，建设门槛将逐步降低，经济可行性会越来越高。

5.2.2 推广可行性

推广存在以下主要驱动力：

(1) 巨大的效能诱惑，以“效率提升 10-100 倍”为最核心的推广点。一旦有成功案例，会形成强大的示范效应和竞争压力。

(2) 数据驱动决策的文化，随着越来越多的组织开始相信数据而非单纯的经验，这为 AI 平台的推广奠定了文化基础。

(3) 解决“专家荒”难题。在特定领域，资深专家稀缺，AI 平台可以提供专业知识，让普通技

术人员也能完成复杂探索。

推广的主要障碍有：

高昂的切换成本，抛弃传统的“博士-实验员-烧杯”模式，转向“AI-机器人-代码”模式，需要彻底的流程再造和组织架构调整。

信任与可解释性弱，化学家如何信任一个“黑箱”给出的建议？缺乏模型可解释性是其被广泛接纳的核心障碍。需要证明AI不仅“有效”，而且“可靠”、“可信”。

存在人才缺口，极度缺乏既懂化学又精通AI和机器人技术的“跨界”人才。大学的教育体系尚未完全跟上时代。

知识产权与数据隐私，企业担心其核心的化学数据在上云或使用外部平台时泄露，降低了一部分积极性。

推广意见：

将从“点”到“面”逐步推广，平台不会一夜之间普及。最初会在企业内部，从某个特定项目或研发环节（如先导化合物优化、催化剂筛选）开始试点，成功后再逐步推广到更多部门。

“AI-HI”（人类智能）协作是突破口：推广的重点不应是“取代化学家”，而是“增强化学家”。平台设计成易用的辅助工具，化学家负责提出关键问题、设计探索边界、解读最终结果。

教育和培训是关键：需要大力培养新一代的化学家，并为企业现有的研发人员提供再培训。

5.2.3 预期应用场景

以下是几个最具潜力的落地场景：

(1) 药物发现与开发

场景：靶点识别、苗头化合物发现、先导化合物优化（提高活性、降低毒性）、合成路线优化和设计。

可行性极高。药物研发成本高、周期长、失败率高，是AI最能体现价值的领域。

(2) 功能材料开发

场景：新型电池材料（正极/负极/电解质）、OLED发光材料、高性能催化剂（如用于碳中和的催化剂）、高分子聚合物、金属合金。

可行性极高。材料性能与微观结构、组分关系复杂，非常适合AI进行高通量筛选和优化。

(3) 合成化学与工艺优化

场景：探索新反应、寻找绿色温和的反应条件、优化现有化工过程的产率和纯度。

可行性很高。能够快速遍历巨大的反应条件空间，找到最优解，直接降低生产成本和环境足迹。

(4) 特种化学品与配方产品

场景：开发化妆品、涂料、粘合剂、洗涤剂、香精等复杂配方。

可行性：中到高。配方通常是多组分混合物，其宏观性质与组分间相互作用非线性相关，AI能有效处理这种复杂性。

(5) 基础科学研究

场景：验证化学理论、探索未知的化学空间、发现新的物理化学规律。

可行性很高。作为强大的探索工具，可以产生海量高质量数据，驱动理论化学的进步。

5.3 目标人群

科研人员：包括但不限于高等学校教授和研究生、研究所研究员；

企业人员：包括但不限于科学家，研发工程师，工艺工程师，前沿行业的初创公司；

公职人员：包括但不限于政府与监管机构；

医疗部门：包括但不限于医疗研究的医生、药剂师；

教育部门：包括但不限于老师、学生；

其他人员：包括但不限于工人、农民。

5.4 利润最大化转化

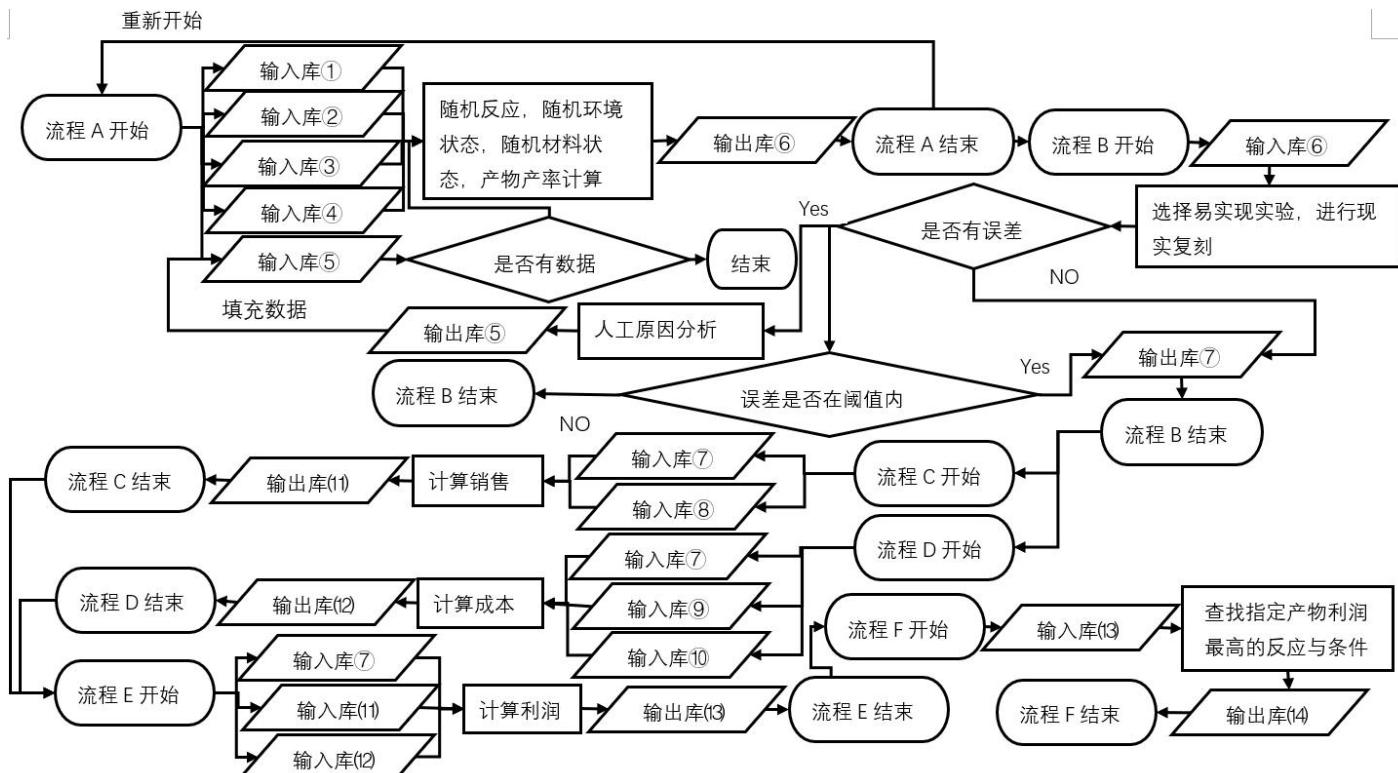


图9：利润转化训练流程

图例和注记

- 库①：化学反应库，含有已知的化学反应
- 库②：材料影响库，含有有不同纯度，不同密度的材料数据以及催化剂数据
- 库③：环境影响库，含有温度压强等因素
- 库④：副反应影响库，含有与反应关联的一些副反应因素
- 库⑤：影响补充库，还有此前未考虑到的影响因素
- 库⑥：产率库，还有此前条件下各个反应产物的产率
- 库⑦：可尝试应用产率库，含有误差在可接受阈值内的反应条件及化学反应和产率数据
- 库⑧：销售价库，含有产物近 1 月内的平均销售价格
- 库⑨：材料成本库，含有反应材料近一个月内的销售价
- 库⑩：环境成本库，含有营造反应所需环境的近一年内的平均价格
- 库⑪：销售额库，含有反应产物变卖所能得到的销售额
- 库⑫：成本库，含有有反应材料及营造反应所需条件的成本
- 库⑬：利润库，含有反应回应条件下的利润
- 库⑭：利润最大库，含有指定产物利润最高的反应与条件

流程叙述

流程 A：产率计算

输入：从「库①（化学反应库）、库②（材料影响库）、库③（环境影响库）、库④（副反应影响库）、库⑤（影响补充库）」导入化学反应、材料特性、环境因素、副反应等基础数据。

处理：结合“随机反应/环境状态、随机材料状态”，计算产物产率。

输出：将产率结果存储至「库⑥（产率库）」，随后流程 A 结束。

流程 B：实验复现与误差验证

输入：从「库⑥（产率库）」导入产率数据，选择易实现的实验进行现实复现。

验证：检查实验是否有误差：

若有误差：进行“人工原因分析”后，再次判断“误差是否在阈值内”；

若误差在阈值内：将“误差可接受的反应条件、化学反应、产率数据”存储至「库⑦（可尝试应用产率库）」；

若无误差：直接进入后续流程。

流程 B 结束后，分支出流程 C、D（并行计算销售额、成本）。

流程 C：销售额计算

输入：从「库⑦（可尝试应用产率库）、库⑧（销售价库）」导入“反应产率、产物近1月平均售价”。

处理：计算“反应产物变卖的销售额”。

输出：将结果存储至「库⑪（销售额库）」，流程C结束。

流程D：成本计算

输入：从「库⑦（可尝试应用产率库）、库⑨（材料成本库）、库⑩（环境成本库）」导入“反应产率、材料近1月成本、环境营造近1年平均成本”。

处理：计算“反应材料+环境条件”的总成本。

输出：将结果存储至「库⑫（成本库）」，流程D结束。

流程E：利润计算

输入：从「库⑦（可尝试应用产率库）、库⑪（销售额库）、库⑫（成本库）」导入“反应产率、销售额、总成本”。

处理：通过「利润=销售额-成本」计算利润。

输出：将结果存储至「库⑬（利润库）」，流程E结束。

流程F：查找利润最高的反应与条件

输入：从「库③（环境影响库）」导入环境因素（辅助筛选条件），结合「库⑬（利润库）」的利润数据。

处理：查找“指定产物利润最高的反应与条件”。

输出：将结果存储至「库⑭（利润最大库）」，流程F结束。

核心逻辑

从“化学反应可行性（产率）”出发，通过实验证可靠，再融入“市场-成本”的经济分析，最终筛选出“利润最大化”的反应及条件，实现“技术可行+经济最优”的双重目标。

5.5 理论研究方面转化方向设想

AIChem 平台在研究理论化学领域上有巨大的潜力，包括但不限于新反应路径与机理推导，设计并预测分子间作用，预测物质性质，化学理论的辅助验证，计算化学与实验的融合，化学暗物质探索等方面，下面将从这六个方面着重论述：

5.5.1 新反应路径与机理推导

传统的机理研究依赖于表征中间体、同位素标记、理论计算等，研究进程缓慢。而 AIChem 平台可以控制自动化反应器，在庞大的反应物、催化剂、溶剂和条件（温度、光照）空间中进行高通量实

验，进而发现那些非直觉的新反应，促进高维反应路径探索。

此外，通过利用实时监测反应（如在线红外、质谱）产生的海量动力学数据，AIChem 平台可以逆向推导出最可能的反应机理网络，甚至找到难以通过实验捕获的瞬态中间体。

最后，AIChem 平台通过学习分子结构与区域/立体选择性之间的深层关系，对于有多个可能产物的复杂反应，可以预测在何种条件下能得到单一的目标产物，从而为合成设计提供理论指导^[42]。

5.5.2 设计并预测分子间作用

AIChem 平台可高通量合成海量的分子组合，并通过自动化 X 射线衍射（XRD）等表征其结构，学习数据后能够仅从分子结构预测两个或多个分子是否能形成稳定的共晶或超分子结构，并预测其晶体结构图谱，做到共晶、超分子组装预测^[43]。

材料科学和药物开发中理解分子在溶液中的聚集，胶束化行为，或在固态下的堆积方式是至关重要的一环。AIChem 平台可以建立从分子结构到其宏观聚集行为的映射，揭示溶液行为与聚集态预测规律。

5.5.3 预测物质性质

平台可以整合多维数据（分子结构、反应条件、谱图数据、生物活性等），构建更强大、更通用的定量结构-性质关系模型，用于预测诸如催化剂活性、材料带隙、溶解度、毒性等复杂性质^[44]。

5.5.4 化学理论辅助验证

AIChem 平台可以通过进行大规模、系统性的实验来验证一个新的化学理论（例如，关于某种催化循环的假设）。平台可以设计并执行一系列“关键实验”，快速证实或证伪该理论。

此外，AIChem 平台在广阔的化学空间中探索时，很可能会发现大量与现有理论预测不符的“异常值”，进而发现理论中的例外与边界。它们是推动理论修正和突破的宝贵契机。

最后，通过对海量实验数据进行无监督或半监督学习，AIChem 平台可能识别出人类尚未注意到的、支配化学行为的新的经验规则或描述符，从而为理论化学家提供构建新理论的线索，提出数据驱动的理论。

5.5.5 计算化学与实验的融合

AIChem 平台可以快速生成大量、高质量的实验数据，为不同层次的理论计算方法（如 DFT、MD）提供一个强大的基准测试集，评估哪种方法在预测特定性质时最准确。

此外，平台提供的实验数据（如光谱、热力学数据）可以用于优化和验证力场参数，使其更接近现实。将高通量实验数据与高通量量子化学计算数据相结合，训练一个代理模型。这个模型可以像高级计算化学方法一样准确，但计算成本低廉到可以瞬间完成预测，从而实现“超尺度”的化学空间筛

选^[45]。

5.5.6 化学暗物质探索

当前人类所知的化学是有限的。AIChem 平台可以作为一个自主的探索者，系统性地探索那些人类从未涉足、甚至从未想象过的化学区域，发现具有奇异性质和全新结构模式的“化学暗物质”^[46]。

5.6 工程实践方面转化方向设想

AIChem 平台在诸多领域都有着巨大的应用潜力，包括但不限于生物医药、新型材料、化工生产、目标物质获取和新医学方向。下面将着重对上述五个方向进行说明。

5.6.1 药物开发领域

药物研发领域面对着疾病机制的复杂性与靶点的不可成药性等挑战。需要对大量化合物进行筛选，但命中率极低，筛选物往往活性选择性或药性不佳，优化方面难以满足全部需求。动物实验与模型具有局限性^[47]。

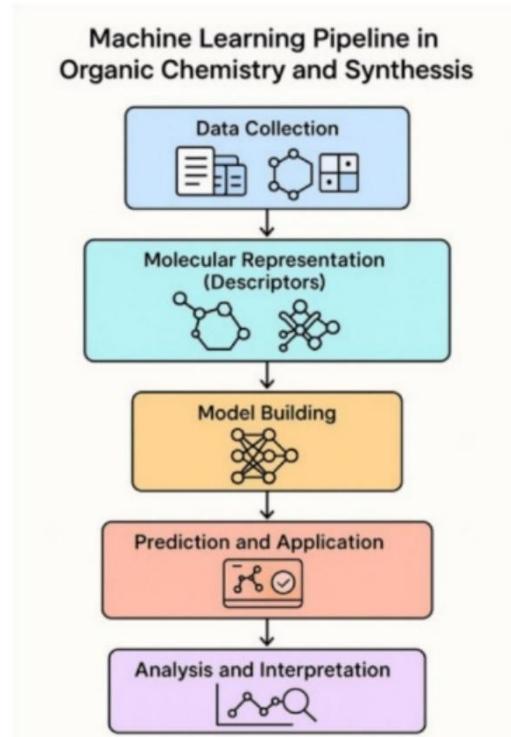


图 10：有机化学合成中机器学习的应用

AIChem 平台可以快速大量地模拟出尚未在实验室进行合成的有机物的理化性质，其中部分有机物可以作开发新药或新的药物转运载体之用。我们可以将此项技术与生理药代动力学模型进行结合，通过将模拟出的理化数据导入到计算机的生理药代动力学模型中，通过推算得到 AIChem 平台模拟出的药物在动物体中的药效，这样可以大幅度减少人力工作与在实验室进行实验的次数。

5.6.2 材料领域

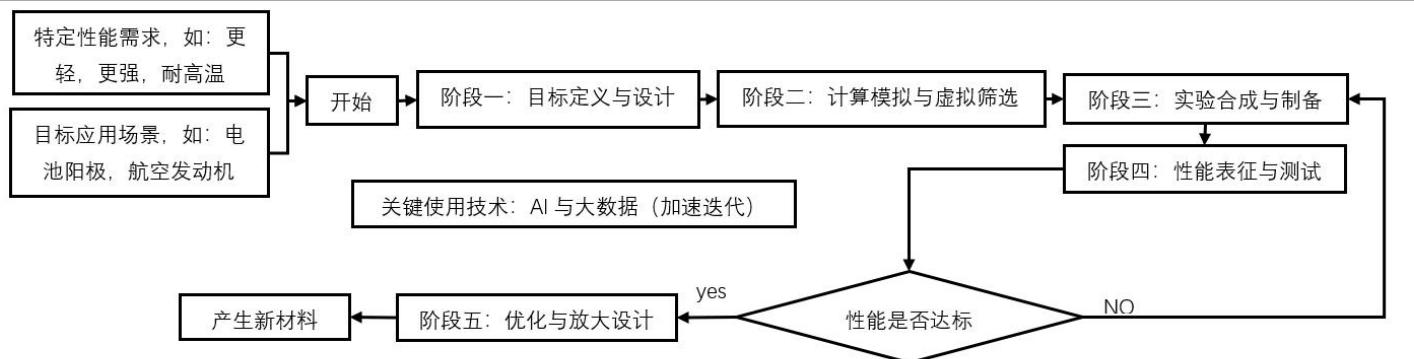


图 13：新材料研发专门化流程

新型材料领域面临的核心困境是“研发与应用之间的高墙”。研发周期极长且成本高昂，从实验室的理论构想、合成制备到性能优化，需要经历无数次的试错与迭代。性能与工艺的平衡难题。许多在实验室表现优异的材料，一旦进入规模化生产，常面临窘境。最后，环保与可持续发展压力日益增大。这些因素共同构成了制约新材料从实验室走向市场的巨大瓶颈^[48]。

AIChem 平台还可以模拟出一些在实验室中尚未合成的新型材料的理化数据，这些数据同样可以通过与其他理论模型结合来预测这些未合成的材料的性能。通过将这些数据在经典分子动力学模型中进行模拟，我们可以得到在人工智能化学平台中推演出的新型材料的性能数据，这样的结合可以加快新型材料的研发探索。

5.6.3 形成全新的化工生产方式（由 AI 主导的化工简易路线）

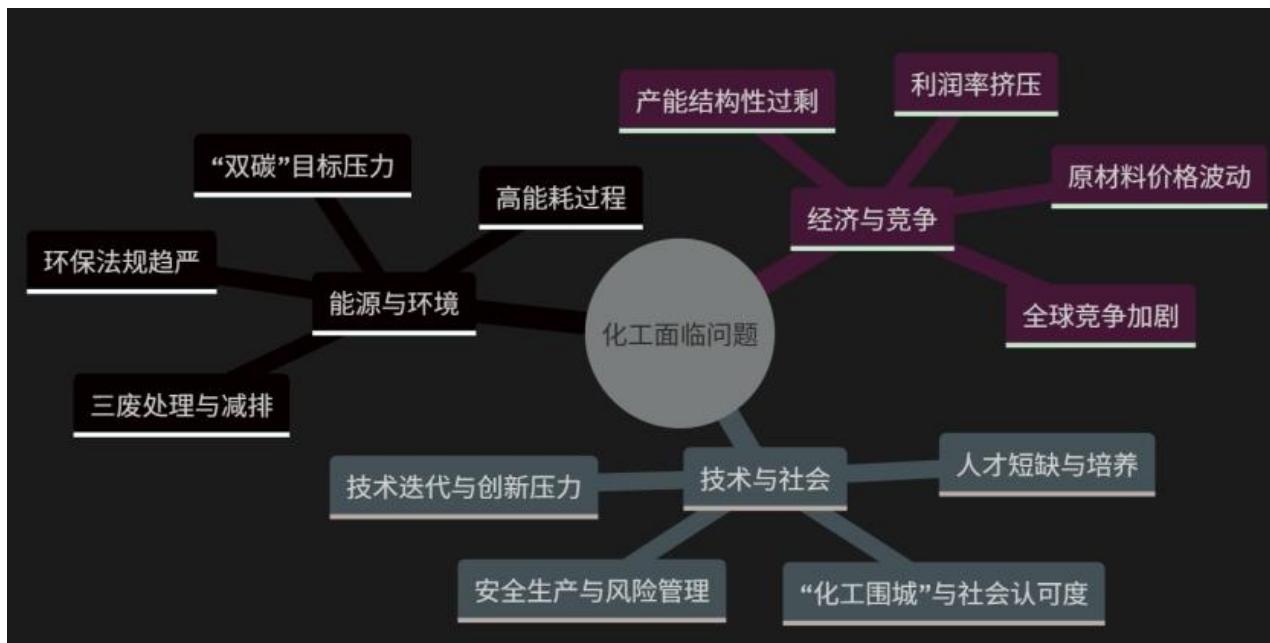


图 14：化工生产困境

化工领域面临着生产材料成本过高，排放不达标，利润率挤压等问题，该技术可以扩大生产材料

的方向，找出成本低的化工原料，以及安全性高，低污染甚至无污染的化工流程。

通过分析海量分子数据与反应路径，AIChem 平台能快速筛选潜在新材料，精准预测性能与合成路线，显著缩短研发周期。它不仅能优化现有工艺，更能发现传统方法难以触及的高价值生物基、废弃资源转化材料，为化工行业开辟降本增效、绿色可持续的新原料来源，驱动产业创新。

5.6.4 获取目标物质

人类目前在获取新物质方面存在以下问题，其一，数据瓶颈，高质量、大规模的实验数据稀缺；其二，模型“黑箱”特性导致预测结果难以解释和信任；其三，对复杂化学反应机理、热力学稳定性及实际合成路径的精准模拟仍是巨大难题，限制了其从虚拟预测走向现实应用的步伐。

通过将自编码器（autoencoder，AE，是一类在半监督学习和非监督学习中使用的人工神经网络（Artificial Neural Networks，ANNs），其功能是通过将输入信息作为学习目标，对输入信息进行表征学习）融入到此人工智能化学平台的应用端中，输入此平台的数据让融入的人工神经网络进行学习，我们可以开发出此平台的一个全新应用：向平台输入目标要求的理化性质，通过平台模拟出可能性最大的物质结构。这项应用可以大幅提高开发新型物质工作的效率^[49]。



图 15：未来 AI 分析物质假想图

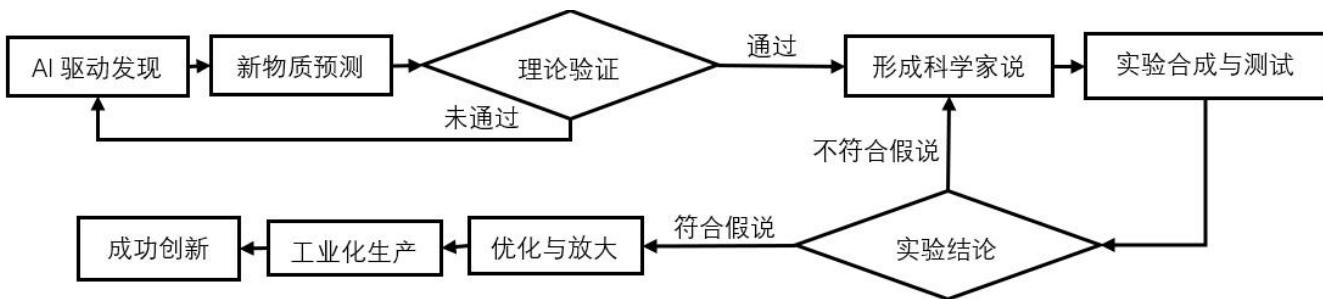


图 16：目标物质获取专门化流程

5.6.5 现代中医药研究相结合

人工智能正重塑中医药研发范式。通过深度学习古籍方剂与临床数据，AI能高效挖掘药效成分、解析“君臣佐使”配伍规律，并预测药物作用机制。同时，借助虚拟筛选与分子生成技术，加速发现新型化合物，显著提升研发效率。此外，AI赋能的中医智能辅助系统，通过数据化“望闻问切”，推动诊疗的标准化与精准化，为这一传统领域注入现代科技动能^[50]。

通过让 AIChem 平台对大量中药中重要的有效成分的化学结构进行学习，我们不仅可以通过将平台与 ML、DL 算法相结合，将平台用于分子层面数智中药的研发，还可以通过平台获得的自主学习能力帮助进行中药中天然产物的虚拟筛选。

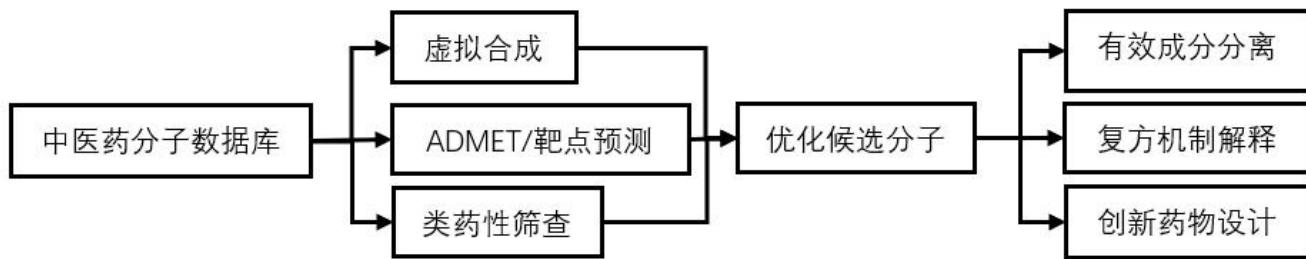


图 17：新型医药研发专门化流程

参考文献：

- [1] Zhang Y, Han Y, Chen S, et al. Large language models to accelerate organic chemistry synthesis[J]. Nature Machine Intelligence, 2025, 7: 1010-1022.
- [2] Zheng M Y, et al. Bridging chemistry and artificial intelligence by a reaction description language[J]. Nature Machine Intelligence, 2025.
- [3] 房喻. 人工智能赋能教育：从跟随 ChatGPT-4 进行的一次自主学习说起[J]. 大学化学, 2025, 40(X): 1-3. DOI:10.12461/PKU.DXHX202502013.
- [4] 沈湫莎. AI 赋能化学，机器人如何“灵光一现” [N]. 文汇报, 2025-07-06(002).
- [5] 沙明瑞. 人工智能算法在化学反应预测中的应用[J]. 当代化工, 2025, 54(7): 1710-1715. DOI:10.13840/j.cnki.cn21-1457/tq.2025.07.024.
- [6] 清华大学团队.《SurFF: A Foundation Model for Surface Exposure and Morphology Across Intermetallic Crystals》 [J].Nature Computational Science, 2025. DOI: 10.1038/s43588-025-00912-1. <https://www.nature.com/articles/s43588-025-00912-1>
- [7] 清华大学团队.《Theoretical High-Throughput Screening of Single-Atom CO₂ Electroreduction Catalysts to Methanol Using Active Learning》 [J].Engineering, 2025. DOI: 10.1016/j.eng.2025.08.005. <https://www.sciencedirect.com/science/article/pii/S2095809925003026>

- [8]清华大学团队.《CaTS: Toward Scalable and Efficient Transition State Screening for Catalyst Discovery》[J].*ACS Catalysis*, 2025. DOI: 10.1021/acscatal.5b03478. <https://pubs.acs.org/doi/10.1021/acscatal.5b03478>
- [9]清华大学团队.《Bayesian-Learning-Assisted Catalyst Discovery for Efficient Iridium Utilization in Electrochemical Water Splitting》[J].*Science Advances*, 2025. DOI: 10.1126/sciadv.adg5234. <https://www.science.org/doi/10.1126/sciadv.adg5234>
- [10] 佚名. 机器人能用AI加速发现化学分子[N]. 科技日报, 2024-01-26(信息与动态). [2024-02-15].
- [11] 张梦然. 两台AI机器人联手开展化学研究 未来科研方式或将改变[N]. 科技日报, 2024-11-11(004).
- [12] The Nobel Foundation. Nobel Prize in Chemistry 2024[EB/OL]. (2024-10-09 15:53). <https://www.nobelprize.org/prizes/chemistry/2024/summary/>.
- [13] Okada M. 2024 Nobel prizes in physics and chemistry: from neural network models to materials engineering[J]. *Science and Technology of Advanced Materials: Methods*, 2025, 5(1): 2516307. DOI:10.1080/27660400.2025.2516307.
- [14] 刘霞. 他们用AI破译蛋白质结构“密码”——三位科学家分享2024年诺贝尔化学奖[EB/OL]. (2024-10-09). https://www.stdaily.com/web/gdxw/2024-10/09/content_240160.html.
- [15]洛斯阿拉莫斯国家实验室团队.《Breakthrough Reported in Machine Learning-Enhanced Quantum Chemistry》[J].*Science*, 2025. DOI: 10.1126/science.adh3456. <https://www.science.gov/topicpages/b/breakthrough+reported+in+machine.html> [13]MIT团队.《High-Throughput Discovery of Ferrocene Mechanophores with Enhanced Reactivity and Network Toughening》[J].*ACS Central Science*, 2025. DOI: 10.1021/acscentsci.5b00789. <https://pubs.acs.org/doi/10.1021/acscentsci.5b00789>
- [16]密歇根大学团队.《Learning Local and Semi-Local Density Functionals from Exact Exchange-Correlation Potentials and Energies》[J].*Science*, 2025. DOI: 10.1126/science.adh3456. <https://www.science.org/doi/10.1126/science.adh3456>
- [17]南加州大学团队.《Allegro-FM: Toward an Equivariant Foundation Model for Exascale Molecular Dynamics Simulations》[J].*The Journal of Physical Chemistry Letters*, 2025. DOI: 10.1021/acs.jpclett.5b01345. <https://pubs.acs.org/doi/10.1021/acs.jpclett.5b01345>
- [18] Shilpa S, Kashyap G, Sunoj R B. Recent applications of machine learning in molecular property and chemical reaction outcome predictions[J]. *The Journal of Physical Chemistry A*, 2023, 127: 8253-8271. DOI:10.1021/acs.jpca.3c04779.
- [19] Zekić A. Mathematical optimization in machine learning for computational chemistry[J]. *Computation*,

- 2025, 13(7): 169. DOI:10.3390/computation13070169.
- [20] Mahjour B, Shen Y, Cernak T. Ultrahigh-throughput experimentation for information-rich chemical synthesis[J]. Accounts of Chemical Research, 2021, 54(10): 2337-2346. DOI:10.1021/acs.accounts.1c00119.
- [21] Nsouli R, Galiyan G, Ackerman-Biegasiewicz L K G. Advancing organic chemistry using high-throughput experimentation[J]. Angewandte Chemie International Edition, 2025. Article Number: e202506588. DOI:10.1002/anie.202506588.
- [22] Boge F J, Poznic M. Machine learning and the future of scientific explanation[J]. Journal for General Philosophy of Science, 2021, 52(1): 171-176. DOI:10.1007/s10838-020-09537-z.
- [23] Aboy M, Lath A, Minssen T, et al. The sufficiency of disclosure of AI inventions[J]. Journal of Intellectual Property Law & Practice, 2024, 19(11): 834-840. DOI:10.1093/jiplp/jpae063.
- [24] Haynes W M. CRC Handbook of Chemistry and Physics[M]. 104th ed. Boca Raton: CRC Press, 2023.
- [25] 武汉大学, 吉林大学, 南开大学, 等. 无机化学[M]. 第5版. 北京: 高等教育出版社, 2006.
- [26] 北京大学化学与分子工程学院《普通化学原理》编写组. 普通化学原理[M]. 第4版. 北京: 高等教育出版社, 2019.
- [27] Chemeurope. Electronegativities of the elements (data page)[EB/OL]. (2018-06-14). https://www.chemeurope.com/en/encyclopedia/Electronegativities_of_the_elements.html.
- [28] 《中国大百科全书》编辑部. 电负性[EB/OL]. (2024-09-27). <https://www.zgbk.com/ecph/words?SiteID=1&ID=437827&Type=bkzyb&SubID=112425>.
- [29] Chem.libretexts.org. 2.5C: Allred-Rochow electronegativity values[EB/OL]. (2023-01-16). [https://chem.libretexts.org/Bookshelves/General_Chemistry/Chem1_\(Lower\)/02%3A_Atoms_Molecules_and_Ions/2.05%3A_Electronegativity_and_Polarity/2.5C%3A_Allred-Rochow_Electronegativity_Values](https://chem.libretexts.org/Bookshelves/General_Chemistry/Chem1_(Lower)/02%3A_Atoms_Molecules_and_Ions/2.05%3A_Electronegativity_and_Polarity/2.5C%3A_Allred-Rochow_Electronegativity_Values).
- [30] Schoolbag.info. 11.2 Electronegativity[EB/OL]. (无发布时间). <https://schoolbag.info/chemistry/11/112.html>.
- [31] Chem.libretexts.org. 8.8.2: Table of electronegativities[EB/OL]. (无发布时间). [https://chem.libretexts.org/Bookshelves/Inorganic_Chemistry/Map%3A_Inorganic_Chemistry_\(Housecroft\)/08%3A_Basic_Concepts_in_Bonding/8.08%3A_Electronegativity/8.8.2%3A_Table_of_Electronegativities](https://chem.libretexts.org/Bookshelves/Inorganic_Chemistry/Map%3A_Inorganic_Chemistry_(Housecroft)/08%3A_Basic_Concepts_in_Bonding/8.08%3A_Electronegativity/8.8.2%3A_Table_of_Electronegativities).
- [32] IUPAC. XML-based IUPAC standard for experimental, predicted, and critically evaluated thermodynamic property data storage and capture (ThermoML) (IUPAC Recommendations 2006)[J]. Pure and Applied Chemistry, 2006, 78(3): 541-612.
- [33] IUPAC. Extension of ThermoML: the IUPAC standard for thermodynamic data communications (IUPAC Recommendations 2011)[J]. Pure and Applied Chemistry, 2011, 83(10): 1937-1967.
- [34] IUPAC. ThermoML - an XML-based IUPAC standard for storage and exchange of experimental

- thermophysical and thermochemical property data[EB/OL]. (2019-01-05).
<https://iupac.org/project/2017-043-1-022/>.
- [35] 中国大百科全书出版社编辑部. 中国大百科全书·化学[M]. 北京: 中国大百科全书出版社, 1992. ISBN 9787500002239.
- [36] Lide J (Ed.). CRC Handbook of Chemistry and Physics[M]. 105th ed. Boca Raton: CRC Press, 2024: 1604. ISBN 978-1032655628.
- [37] Jisuanla. Advanced standard temperature and pressure (STP) calculator[EB/OL]. (2025-07-17).
<https://www.jisuanla.com/stp-calculator>.
- [38] Dean J A. Lange's Handbook of Chemistry[M]. 14th ed. New York: McGraw-Hill, 1992.
- [39] Lide D R. CRC Handbook of Chemistry and Physics[M]. 77th ed. Boca Raton: CRC Press, 1996.
- [40] 印永嘉. 物理化学简明手册[M]. 北京: 高等教育出版社, 1988.
- [41] 卢焕章. 石油化工基础数据手册[M]. 北京: 化学工业出版社, 1982.
- [42] Ha T, Lee D, Kwon Y, et al. AI-driven robotic chemist for autonomous synthesis of organic molecules[J]. Science Advances, 2023, 9(44): eadj0461. DOI:10.1126/sciadv.adj0461.
- [43] Jin W G, Coley C W, Barzilay R, et al. Predicting organic reaction outcomes with Weisfeiler-Lehman network[C]//Advances in Neural Information Processing Systems 30 (NIPS 2017). Red Hook: Curran Associates, 2017.
- [44] Tamega G S, Costa M O, Pereira A D, et al. Data science guiding analysis of organic reaction mechanism and prediction[J]. Chemical Record, 2024, 24(12). DOI:10.1002/tcr.202400148.
- [45] Coley C W, Barzilay R, Jaakkola T S, et al. Prediction of organic reaction outcomes using machine learning[J]. ACS Central Science, 2017, 3(5): 434-443. DOI:10.1021/acscentsci.7b00064.
- [46] Abolhasani M, Kumacheva E. The rise of self-driving labs in chemical and materials sciences[J]. Nature Synthesis, 2023, 2(6): 483-492. DOI:10.1038/s44160-022-00231-0.
- [47] Francis C P Law (罗松柏), 何绍雄. 生理药代动力学模型的进展与应用[J]. 药学学报, 1997, 32(2): 151-160.
- [48] 樊康旗, 贾建援. 经典分子动力学模拟的主要技术[J]. 微纳电子技术, 2005(3): 133-138.
- [49] 唐楠楠, 张姝, 李强林. 密度泛函理论的基本计算方法研究进展[J]. 成都纺织高等专科学校学报, 2015, 32(2): 39-43.
- [50] 郭鹏飞, 魏宇, 杨鹏程, 等. 人工智能赋能的现代中药数智化研究进展[J]. 中国医药科学, 2025(4): 20-40. DOI:10.3969/j.issn.1673-5390.2025.04.00.

- 备注：**
1. 本申报表须由申报者或团队负责人本人填写；
 2. 团队参赛的须确定负责人；
 3. 正文内容表述要求：宋体小四，1.5 倍行距，英文 Times New Roman；
 4. 本表由大赛组委会负责最终解释。