AI+材料基因工程创新平台

1. 数据获取（实验、收集、采集、整理、筛选、清洗）

材料基因工程的核心在于高质量、多维度的材料数据资源。本研究将围绕高分子电子胶材料，从已有文献、实验报告、专利数据库、商业数据库以及实际实验中获取大量原始数据，涵盖材料组分、结构特征、制备参数和性能指标等。针对原始数据存在的多源异构、格式不统一、缺失值等问题，制定标准化数据处理流程，包括数据清洗、去重、归一化、异常值检测与补全等技术手段，确保数据的完整性和准确性。同时，借助自动化爬虫、自然语言处理等工具高效提取潜在数据源，构建可扩展的数据收集管道。通过系统性整理和筛选，形成结构清晰、信息完备、便于建模的高质量材料数据集，为后续数据库建设和AI建模提供坚实的数据基础。

2. 构建高分子电子胶材料的组分-结构-性能数据库

在完成数据获取与清洗的基础上，构建高分子电子胶材料的“组分-结构-性能”三元数据库。该数据库将材料的化学组成、微观结构（如分子链构型、交联密度、结晶度）、工艺条件与宏观性能（如电导率、粘接强度、柔韧性等）进行关联，形成结构化、层次化、多尺度的数据体系。采用本体建模和语义标注技术，提升数据库的知识表达能力和可检索性，支持不同层级材料特征的精准查询与分析。同时设计开放式数据库架构，便于未来集成多种材料类型和数据来源。通过图数据库、知识图谱等新型数据结构提升信息的关联性与智能搜索能力，使该数据库不仅具备存储功能，更能作为材料设计与性能预测的重要知识支撑平台。

3. 基于小样本机器学习，构建高分子电子胶材料组分-结构-性能的AI模型

考虑到实验数据获取成本高、数量有限的特点，本研究采用小样本学习方法构建材料组分-结构-性能映射的AI预测模型。结合迁移学习、元学习和自监督预训练等技术，在已有的材料大模型基础上，针对电子胶特性进行微调优化，实现小数据条件下的高精度建模。模型输入包括分子式、SMILES字符串、图神经网络提取的结构特征、RDKit分子描述符、工艺参数等多模态特征；输出则为各项关键性能指标。采用集成学习与不确定性估计方法提升模型鲁棒性与解释性，同时设计可视化分析模块，帮助研究人员理解不同组分与结构对性能的影响机制。该AI模型将为新材料筛选与结构优化提供科学预测依据。

4. 实现新型电子胶材料的AI推荐、快速验证与实验验证

基于构建的AI模型与数据库，建立新型高分子电子胶材料的智能推荐系统，实现从组分设计到性能预测的一体化流程。系统可根据目标性能指标（如导电性、热稳定性、柔性等），自动生成一系列候选材料组合，供科研人员优选。为提高效率，设计快速计算验证机制，如分子动力学模拟、密度泛函理论（DFT）计算等，初步筛除不符合性能预期的材料，减少实验试错成本。最终，将优选材料方案进行实际实验验证，反馈实验数据至数据库，形成“AI推荐-仿真验证-实验验证-数据反馈”闭环机制。通过持续优化与数据迭代，实现新型高性能电子胶材料的高效发现与快速应用，加快材料从研发到产业化的转化进程。