

# 重庆大学化学化工学院 AI 化学与智能创造微专业

## 《化学人工智能方法与应用》

### 课程教学大纲

一、课程名称： 化学人工智能方法与应用

二、学时和学分： 32 学时，2 学分

三、授课对象： 面向全校理工医类专业本科三年级及以上学生开放，重点支持化学与材料、化工与制药、生物与环境、计算机与自动化等方向学生开展“模型—流程—设备—数据”一体化的自动化实验与系统工程交叉学习；对具体仪器与控制经验不作硬性要求，但建议具备基本编程与数据处理素养。

四、先修课程：

必须具备 Python 编程基础，了解基本的化学数据结构（如 SMILES、分子储存格式）。建议具备高等数学（微积分、线性代数）基础知识。

五、开课单位： 化学化工学院

六、授课教师：

序号	教师姓名	工号	职称	所在单位
1	孙耿	30009203	副教授	化学化工学院
2	陈书森	33230	副教授	化学化工学院

七、课程描述：

本课程是数智化学微专业的“智慧大脑”，旨在讲授如何利用人工智能算法解决复杂的化学问题。课程内容贯穿了从经典机器学习到前沿大模型的完整技术谱系。

课程首先引入 AI 在化学中的基本范式，并深入讲解随机森林、XGBoost 等传统机器学习算法在分子性质预测中的应用。随后讲授深度学习前沿，重点包括用于分子结构建模的图神经网络（GNN）、用于分子逆向设计的生成模型（VAE/GAN/Diffusion），以及最新的大语言模型（LLM）在化学文献分析与智能体（Agent）工具调用中的应用。

通过本课程，学生将完成从“使用工具”到“设计模型”的转变，掌握构建化学预测模型、生成新颖分子结构以及开发化学智能助手的核心能力。

表 1 教学环节对应的教学内容

教学环节	教学内容
化学 AI 问题范式与数据到模型映射	化学问题的数学抽象、回归/分类/生成/表示学习四大范式、模型泛化能力的本质、化学任务的建模转化
传统机器学习方法	分子指纹与特征工程、支持向量机 (SVM) 与核函数、集成学习算法 (Random Forest, XGBoost)、超参数调优策略；性质预测与条件分类
表示学习与分子图建模	深度学习基础与反向传播、图神经网络 (GNN) 核心架构、消息传递机制 (Message Passing)、分子图嵌入与端到端预测
生成模型与分子设计	变分自编码器 (VAE) 与潜在空间映射、生成对抗网络 (GAN)、扩散模型 (Diffusion Model)、分子逆向设计与多目标优化
大语言模型与化学推理	Transformer 架构简介、提示工程 (Prompt Engineering)、化学文献摘要与知识抽取、知识抽取、实验规划、工具调用
模型评估与综合项目	误差分析指标 (MSE)、过拟合与欠拟合、模型可解释性分析；全流程综合建模项目实战

## 八、教学目标

教学环节	知识贡献	能力贡献	素质贡献
化学 AI 问题范式	理解化学问题如何转化为计算任务，掌握 AI 解决科学问题的底层逻辑	能够准确界定科学问题类型（回归/分类/生成），选择正确的技术路线	建立计算思维，形成“数据驱动科学发现”的方法论意识
传统机器学习	掌握经典 ML 算法的数学原理与适用场景，理解特征工程的重要性	能够熟练使用 Scikit-learn 进行模型训练与调优，解决小样本化学数据问题	培养工程实践能力，在算法选择上建立“奥卡姆剃刀”式的务实态度
表示学习与 GNN	理解端到端学习与图数据的消息传递机制，突破传统描述符的局限	能够编写 GNN 代码处理非欧几里得数据（分子图），构建高精度预测模型	强化前沿探索意识，理解深度学习在提取复杂化学特征方面的优势
生成模型与设计	理解生成式模型的概	能够利用生成模型创	激发创新创造思维，体验

	率分布拟合原理，掌握分子逆向设计的核心算法	造训练集中不存在的新分子，并进行性质优化	AI 辅助下的“无中生有”分子设计过程
大语言模型与综合	掌握 LLM 的推理能力边界与工具调用原理，理解模型评估的严谨性	能够开发基于 LLM 的化学智能体，并科学评估模型的 OOD 泛化能力	树立批判性思维，警惕 AI“幻觉”，关注 AI 在科学伦理与安全性上的边界

## 九、教学方法：

**分阶段教学：**前半部分打牢经典机器学习基础，后半部分带领学生进入深度学习与大模型的前沿领域。

**实践代码为主：**理论课紧随 Colab/Notebook 实战，强调“所学即所写”，让学生亲手实现从 RF 到 GNN 的每一行核心代码。

**综合项目驱动：**最后 6 学时通过一个完整的综合项目，串联起数据清洗（课程 2）、模型训练（课程 4 前半段）与分子生成（课程 4 后半段），模拟真实的 AI for Science 研究流程。

## 十、考核及成绩评定方式：

**平时作业与实验报告（30%）：**各章节的代码练习与实验结果分析。

**阶段性小测（20%）：**关于算法原理（如反向传播推导、SVM 损失函数）的理论测试。

**期末综合项目（50%）：**

**项目代码（20%）：**代码的可运行性、规范性及模型性能。

**项目报告（15%）：**问题定义的清晰度、方法选择的合理性、结果分析的深度。

**答辩展示（15%）：**逻辑表达能力及对模型局限性的思考。

## 十一、教材/参考书目：

中文参考教材

1. 《机器学习》，周志华 著，清华大学出版社。
2. 《动手学深度学习》（PyTorch 版），阿斯顿·张、李沐 等著，人民邮电出版社。
3. 《图神经网络：基础、前沿与应用》，吴凌飞 等著，机械工业出版社。

英文参考教材与资源

1. Bharath Ramsundar et al., *Deep Learning for the Life Sciences*, O'Reilly Media.
2. Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, O'Reilly Media.
3. William L. Hamilton, *Graph Representation Learning*, Morgan & Claypool Publishers.