

## 人工智能与化学实验课程建设

杜静<sup>1,2,3</sup>, 于曦<sup>3</sup>, 马骁飞<sup>1,2,3</sup>, 赵温涛<sup>1,2,3,\*</sup>

<sup>1</sup>天津大学化学化工国家级实验教学示范中心, 天津 300072

<sup>2</sup>天津大学化学化工国家级虚拟仿真实验教学中心, 天津 300072

<sup>3</sup>天津大学理学院化学系, 天津 300072

**摘要:** 本文结合人工智能技术及其在化学相关研究中的实际应用, 探讨了人工智能与化学课程的建设方案。提出了人工智能与化学实验课程的教学目标、课程设置与教学大纲、实施方法及教学效果, 并介绍了人工智能与化学实验课程及AI+化学化工虚拟仿真平台建设。为我国化学教育及未来化学研究提供助力。

**关键词:** 人工智能与化学; 化学信息学; AI+化学化工虚拟仿真平台

**中图分类号:** G64; O6

## Artificial Intelligence & Chemistry Course Construction

Jing Du<sup>1,2,3</sup>, Xi Yu<sup>3</sup>, Xiaofei Ma<sup>1,2,3</sup>, Wentao Zhao<sup>1,2,3,\*</sup>

<sup>1</sup> National Demonstration Center for Chemistry and Chemical Engineering Education, Tianjin University, Tianjin 300072, China.

<sup>2</sup> National Virtual Simulation Experimental Teaching Center of Chemical and Chemical Engineering, Tianjin University, Tianjin 300072, China.

<sup>3</sup> Department of Chemistry, Tianjin University, Tianjin 300072, China.

**Abstract:** This paper explores the integration of artificial intelligence technology into the construction of chemistry courses, highlighting its practical applications in chemical research. It outlines the teaching objectives, curriculum design, and syllabus for AI-enhanced chemistry courses, along with implementation strategies and case studies. Additionally, the paper introduces the development of an AI+chemical engineering virtual simulation platform. The establishment of these courses aims to significantly enhance chemical education and research in China.

**Key Words:** Artificial intelligence and chemistry; Chemoinformatics;  
AI+Chemical engineering virtual simulation platform

随着世界范围内新一轮科技革命和产业变革的加速进行, 以新技术、新业态、新产业为特点的新经济获得了蓬勃发展, 互联网、云计算、人工智能等新技术也极大地影响着人们的生产生活<sup>[1-3]</sup>。运用机器学习人工智能方法已经成为全球化学研究领域的一个热潮, 取得了令人瞩目的成就<sup>[4-7]</sup>。在《Science》杂志2021年公布的全世界最前沿的125个科学问题中, 把“AI会重新定义化学的未来吗?”作为化学领域的十大问题之一。机器学习可以在大量的化学信息中寻找发现数据间的联系和规律, 帮助化学工作者做出更合理的判断和决策, 加速化学研发过程<sup>[8-10]</sup>。在机器学习方法的加持下, 在分子动力学模拟、量子化学计算、化学合成路径分析、催化剂性能预测、电池材料开发等领域均获

收稿: 2024-03-23; 录用: 2024-05-31; 网络发表: 2024-06-14

\*通讯作者, Email: wentao\_zhao@tju.edu.cn

基金资助: 2023年教育部产学合作协同育人项目(230803132010913, 230703924284711)

得了显著的进步,人工智能正在逐渐改变化学研究的范式,化学研究正在从“实验-理论-计算”的三驾马车时代快速迈入以大数据和机器学习驱动的新时代。

当前,我国化学专业学生通常接受经典化学理论与研究方法教育及实验的系统训练,而缺乏现代计算数学、信息学、计算机及自动化、现代机器学习的相关知识<sup>[1]</sup>。这一状况严重限制了学生接受现代人工智能技术,并将其运用于化学领域,以及未来从事这一新兴领域的可能性。随着全世界人工智能研究的迅猛发展,化学研究正快速走入智能时代,缺乏人工智能相关化学课程的状况将严重限制我国多个涉化学科的发展,甚至有在未来与发达国家差距拉大的风险,因而在涉化类相关专业,如化工、材料、生物、制药、环境等专业中引入人工智能相关课程、普及机器学习方法,是未来化学专业课程建设的一个方向。

本文面向人工智能和数据驱动,探索化学教育的新方向,将化学信息学、人工智能、大数据等产业和技术的前沿进展及行业对人才培养的新要求引入化学人才培养过程,建设人工智能与化学实验课程,形成能够满足数智化时代发展需要的、可共享的教学建设方案。本文基于天津大学化学系现有专业课程,将现代人工智能理论与化学研究的实际应用相结合,提出了人工智能与化学实验课程的教学目标、课程设置与教学大纲,实施方法与教学案例,并介绍了天津大学化学系建立的化学信息数据库与AI+化学化工虚拟仿真平台。本课程的建立将填补这类课程国内外的空白,为我国的化学教育及未来的化学研究发展向世界领先迈进提供助力。

## 1 教学目标

为培养适应数智时代的新工科人才,本课程旨在培养学生具备坚实科学素养与批判性思维,评估AI技术在化学实验中的有效性,辨识潜在的局限性和偏差;以科学精神和科学素养为基础,融合化学实验与机器学习算法,设计创新性研究项目,推动从传统“查文献预测实验”模式到“大数据科学预测实验”模式的转变。

本课程教学总体目标为:使学生初步掌握机器学习的原理、方法及在化学研究中的应用,包括机器学习的基本原理、化学信息的有效表示、常用机器学习算法,并解决化学问题。通过课程学习,学生能够使用人工智能工具进行化学数据的获取、处理和分析。具体教学目标如下:

(1) 学习如何使用人工智能技术解决化学问题,掌握化学信息的有效表示:能解读和掌握如何有效地表示和编码化学信息,包括分子结构、化学反应等。掌握神经网络、决策树、支持向量机等方法,解决化学问题,如分子模拟、催化剂筛选、有机分子构效关系等。培养学生运用数据挖掘算法,发掘化学数据中的隐藏模式和关联,为化学研究提供新的思路和方法。

(2) 掌握使用Python语言进行化学信息处理和机器学习的技能:学生能利用Python语言对分子信息进行读写,构建分子势能面、系统能量计算等,并利用机器学习方法进行化学性质判断,提升化学信息处理和机器学习的技能。

(3) 熟练使用计算化学和人工智能相关软件:能熟练使用Gaussian、Chemdraw、TensorFlow、Anaconda等软件,并掌握其在化学研究中的应用。通过简单编程,利用超算资源和高性能计算机集群进行计算,提交并正确提取计算结果。

(4) 利用人工智能技术进行化学实验的智能化设计:包括实验条件的优化、反应路径的预测等。能通过建立化学模型,对实验结果进行预测,提高实验效率,降低实验成本。深入理解人工智能与化学之间的交叉点,掌握两者融合的基本原理和方法。运用跨学科知识,提出新颖的化学研究问题,设计创新的解决方案,推动化学领域的进步。

## 2 课程建设方案

为了实现上述教学目标,从以下几个方面开展课程建设。

(1) 构建人工智能与化学实验课程知识图谱:对现有的人工智能理论和技术进行深入研究,明

确机器学习核心知识点与化学性质、反应条件、反应机理等知识点之间的关系，形成知识图谱，构建知识网络；

(2) 对有机分子沸点、熔点、蒸发焓、熔化焓、表面张力、密度、压缩因子以及临界性质数据进行收集、整理与清洗工作，并搭建关系型理化性质数据库。构建AI+化学化工虚拟仿真平台，有机整合现有教学资源，搭建便捷智能的运行环境，为人工智能与化学交叉学科教学提供条件。设计利用人工智能技术解决化学问题的实验，让学生掌握如何使用人工智能技术解决化学问题，包括分子模拟、催化剂筛选、有机分子构效关系等；

(3) 通过调研、研讨和交流合作的方式，制定适用于化学大数据与人工智能课程的教学内容与教学大纲，收集整理合适的教学资料，设计制作教学课件及实践教学案例，创建适合于化学专业的人工智能与化学实验课程。课程的建设将为学生提供更加全面、深入的人工智能化学教育，为他们的未来科研和工作打下坚实的基础。

课程建设思路如图1所示，课程使用的教学工具主要包括：硬件(服务器/工作站)、软件(Gaussian/Chemdraw/TensorFlow等)，从基础的机器学习概念出发，逐步深入人工智能及其在化学领域的应用。为了增强课程的普及性和包容性，我们特地将人工智能与化学实验的内容划分为多个层次，以适应不同基础水平的学生需求。针对本科生和研究生，我们设计了差异化的实验训练类型，确保每位学生都能从中获得有效的知识和技能提升。

### 3 课程实施方案

#### 3.1 实施方法

课程建设与实施主要从课程内容、教学工具、教学方法、效果评价及考核方式四个方面进行。

(1) 课程内容建设包括：从初步认识到熟练掌握最后进行实际应用，形成梯度化的课程内容，使得学生们可以更好地理解人工智能在化学领域中的实际应用。在课程专项训练模块中，学生可结合自身能力与兴趣择其一完成，以满足他们不同的学习需求。

(2) 教学条件：依托化学化工国家级虚拟仿真实验教学中心(天津大学)，中心配备了90个机位和两台工作站，一个高性能计算机集群，该集群目前共有8个节点(其中一个为主管理节点，7个为计算节点)，共计192个核，该集群将为计算化学理论计算、数据库运行以及人工智能运算提供硬件与软件支持。

(3) 教学方法：采用混合式教学方式，构建线上资源包括视频、课件、指导说明等，可以帮助学生更好地掌握软件的使用、服务器远程操控、机器学习运行环境、数据库安装等方面的基础知识。学生可以通过线上慕课的方式进行自学，并在线上讨论区解决问题。线下教学主要是知识点串讲，多种手段解决重点难点问题。采用案例式教学方法，分小组完成不同项目，由每组小组长监督项目任务执行情况并研讨。教学方法既注重线上资源的提供，又注重线下教学的实践性和互动性，可使学生更好地理解课程内容，并在实践中掌握相关知识和技能。

(4) 效果评价及考核方式：课程评价方式采用多元化评价方式，总成绩 = 任务完成情况80% + 线上习题10% + 参与讨论次数与课堂表现 10%。注重对学生的综合素质进行评价，包括思维能力、创新能力、团队协作能力等方面的评价。任务完成情况是总成绩的最大组成部分，主要考查学生对课程内容的掌握程度和实践能力。线上习题和参与讨论次数可以帮助学生更好地理解课程内容，并提高他们的学习效果。课堂表现则可以考查学生的积极性和参与度。

#### 3.2 课程内容

课程总32学时，授课12学时，实验20学时。将以学生为中心，重新设计课程，考虑到学生未来发展的需求和专业知识点的涉及程度。人工智能与化学实验课程，以“应用”为主，让学生在源程序的基础上进行修改和学习。将采用由简入繁的教学方式，从简单的读写到分子性质的预测，再到

更高阶的专项训练，帮助学生逐步提高自己的能力。同时，将为不同层次的学生提供不同的选修模块，以满足不同学生的需求和兴趣爱好。

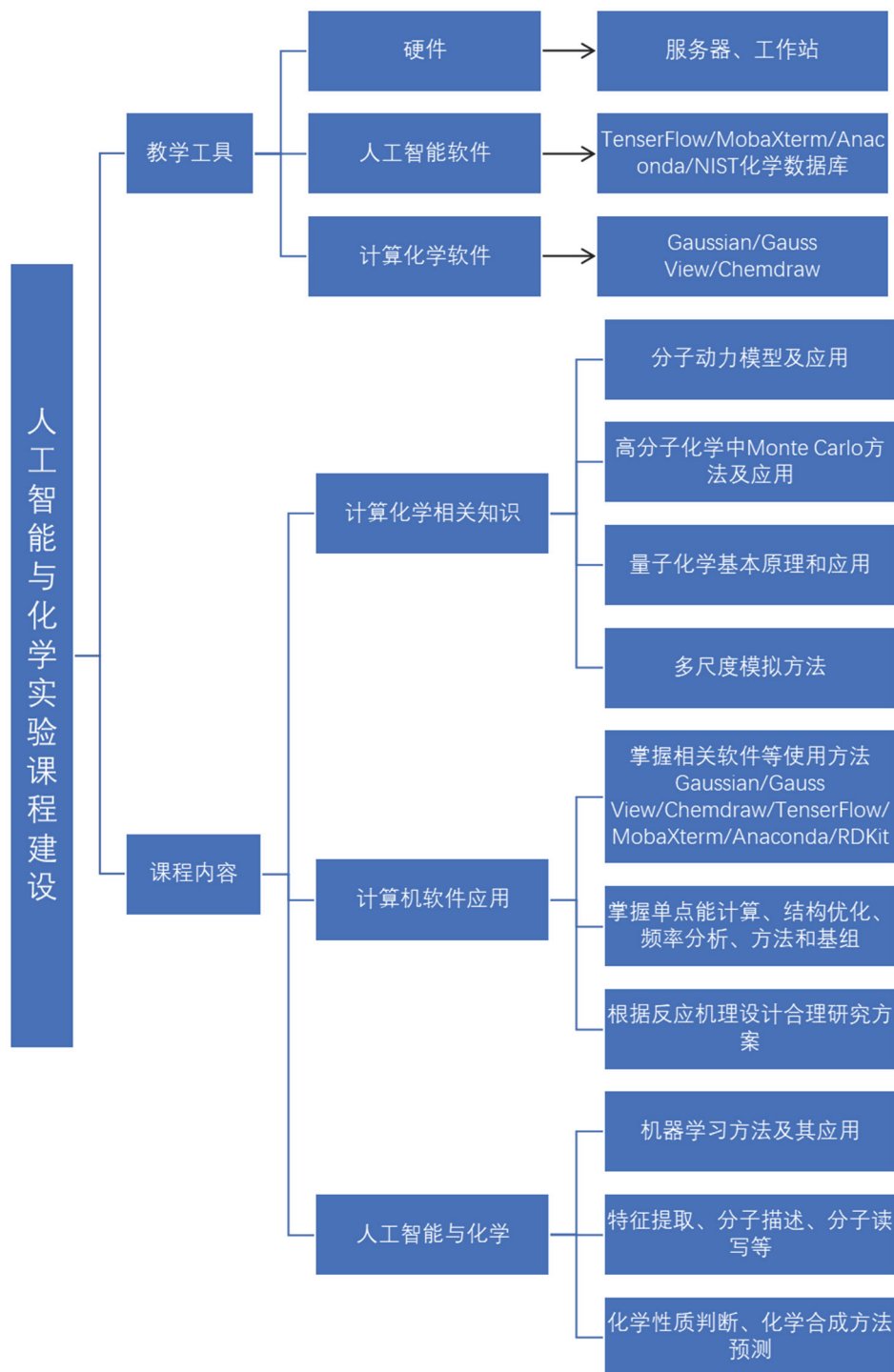


图1 课程建设架构

如表1所示为机器学习与化学实验的课程内容，主要分为基础、提高和专项训练三大模块。在基础层次方面，涵盖机器学习的基础知识、典型机器学习算法的介绍，以及化学分子指纹信

息读写和化学描述符生成的基本能力锻炼。此外，课程还将涉及数据库的搭建和基本化学性质的预测等方面。

在提高层次方面，主要针对研究生，包括构建势能面和神经网络系统训练等方面的内容。通过这些内容的学习，研究生将能够更加深入地理解机器学习在化学领域的应用，并掌握更加高级的技能和知识。

在专项训练方面，由学生自主选择，包括系统能力计算和随机森林机器学习模型的学习等方面。通过这些专项训练，学生将能够更加深入地理解机器学习在化学领域的应用，并掌握更加高级的技能和知识。

表1 人工智能化学课程内容

课程类型	课程大纲	课程内容	授课学时	实验学时
基础(本科)	1 机器学习基础知识	(1) 机器学习的概述 (2) 相关软件应用TensorFlow/MobaXterm/Anaconda/RDKit安装及应用	1	
	2 机器学习算法	(1) 机器学习算法在化学中的应用 (2) 神经网络系统基本架构及编写	2	2
	3 分子指纹	(1) 读写、可视化苯甲酸、乙酰乙酸乙酯、多巴胺三种有机物并提取分子指纹信息以及化学描述符生成 (2) 苯甲酸、乙酰乙酸乙酯、多巴胺三种有机物分子相似性计算	2	2
	4 数据库搭建	(1) 数据库搭建基本方法 (2) 搭建物理化学教材中沸点及凝固点数据库	2	4
	5 化学性质	(1) 多元线性回归函数编写 (2) 未知化学物质催化活性、溶解度判断	2	4
提高(研究生)	6 构建分子势能面	(1) 神经网络系统搭建 (2) 氢分子势能面构建	2	4
	7 神经网络系统	(1) 建立训练集与测试集 (2) 训练两层神经网络系统 (3) 基于机器学习的化合物活性预测		
专项训练(本科+研究生)	8 系统能量计算	(1) 根据反应可能的途径，计算过渡态、中间体的能量 (2) 计算体系反应前后的能量变化	1	4
	9 随机森林	(1) 基于随机森林的机器学习模型预测钾离子通道(hERG)阻断剂活性		

## 4 初步建设成效

### 4.1 应用与实施

目前课程已在应用化学专业大三学生中开课，实施过程中发现处于信息化时代的新一代学生对机器学习相关知识接受程度较高，学习积极性普遍较高，但学生计算机基础参差不齐，学习进度差别较大，可以采用小组制度即由计算机能力较好的学生作为组长帮助组员共同完成教学项目。

图2为学生评价，通过短暂学习让他们对人工智能与化学研究结合有了初步了解，对于涉化类学生学习计算机相关知识是非常必要的，在大数据时代将数据进行科学处理，探索数据中的规律能够为他们今后的科研及走向社会提供更多帮助。学生在人工智能与化学交叉领域获得了一些成绩。在《化学学报》上发表了关于“多组件学习器实现有机分子沸点精准预测”的论文。在第十七届“挑

战杯”中国银行天津市大学生课外学术科技作品竞赛中，获得三等奖。积极参与行业技能竞赛，在第三届DeepModeling Hackathon比赛中获得三等奖，本课程为学生提供了广阔的学习和实践平台。

今天学习了分子指纹+神经网络的程序运行，通过老师的讲解知道了一些计算机语言的相关知识。计算机语言在日常生活中应用广泛，因为现在的工作大多都需要用计算机完成，我们作为当代大学生，是国家未来建设的主力军，我们更要学习好计算机相关方面的知识，并且可以熟练运用计算机，为将来迈入社会参加工作做好准备，也为国家的建设献出自己的一份力。

这次课，老师给我们介绍了人工智能在化学领域中的应用，通过训练机器学习，可以帮助我们完成许多很难的化学实验、计算化合物的分子量、预测化合物的结构等等，在本次实验中我们对人工智能学习与化学研究的结合领域进行了初步的了解。

图2 学生评价

#### 4.2 AI+化学化工虚拟仿真平台建设

如图3(a)所示，AI+化学化工虚拟仿真平台建设主要基于天津大学智能化学团队自建的理化性质预测网站，理化性质预测平台目前具有强大的预测功能，其预测功能主要是基于团队成员构建的数据库信息以及机器学习优化算法，如图3(b)所示为不同理化性质实验数据库，约10万个理化性质数据。基于此获批2023年教育部产学研合作协同育人项目：AI+化学化工虚拟仿真平台建设。

理化性质预测网站如图4所示，目前该网站能够预测包括：沸点、熔点、密度、临界温度、临界压力、临界密度、临界体积、表面张力等有机分子的理化性质，登录后可以通过输入分子标识符或者绘制SMILES进行理化性质预测，对已有的物质性质可以输出准确值，对于未知物质可以输出预测值。

在理化性质预测网站的基础上，添加教学、科研、科普等多维度资源，为相关学者提供学习和交流的平台。通过这个平台，学者们可以共享各自的研究成果和经验，提高彼此的科研水平和能力。同时，平台提供丰富的教学资源，包括课件、视频、案例分析等，帮助学生更好地掌握化学知识和技能。总结并归纳多方面信息与资源为相关学者提供学习交流的平台。

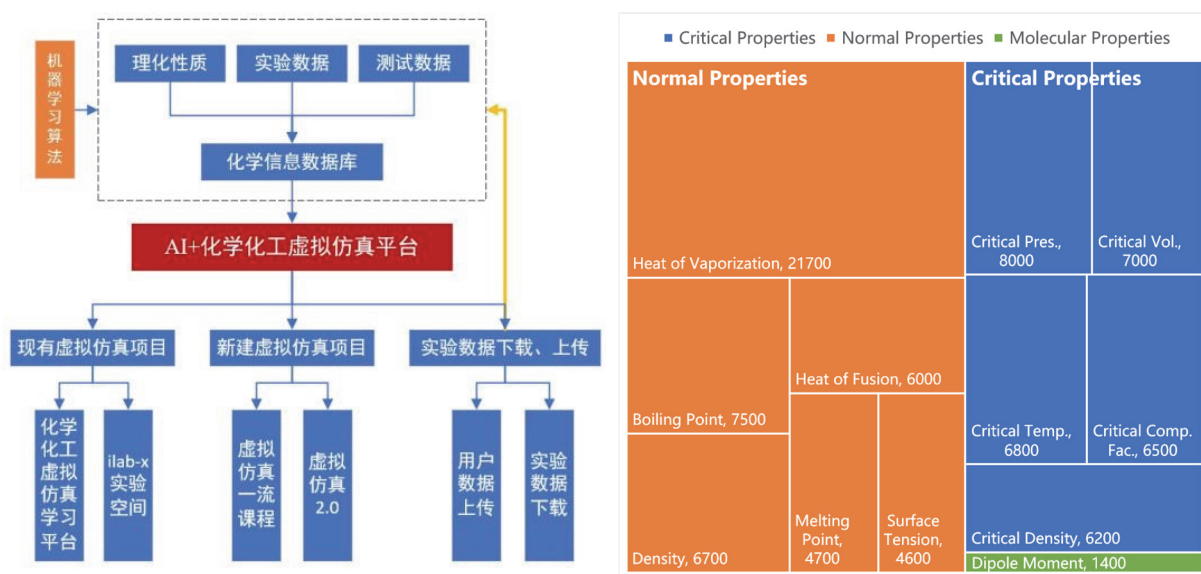


图3 (a) AI+化学化工虚拟仿真平台架构; (b) 不同理化性质实验数据的归纳数目

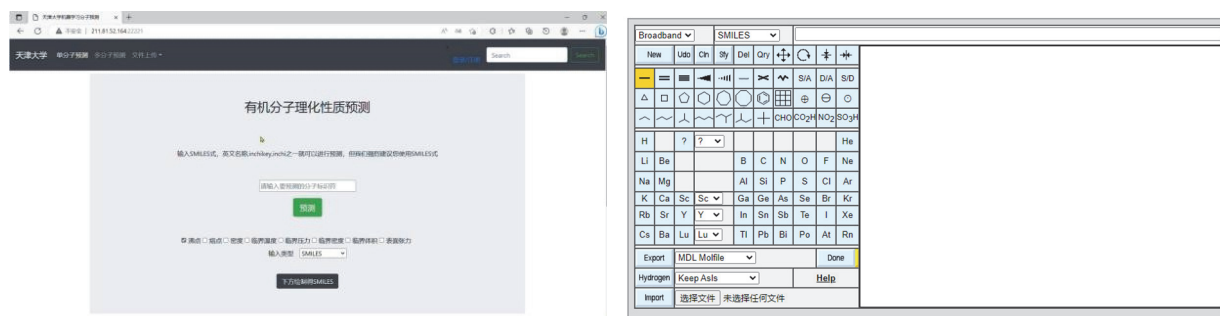


图4 理化性质预测网站

## 5 结语

课程内容结合机器学习在化学领域的实际应用,使学生了解并应用机器学习技术解决化学问题,为未来研究与实践奠定基础。在实际课程实施过程中,结合教学反馈与反思,对教学内容进行迭代。通过课程实施过程中的实践和探索,不断完善和扩展数据库,并将其应用于更加深入和高级的研究领域。课程主要解决以下问题:

(1) 人工智能正在给化学研究带来革命性的变化,国内外尚没有人工智能化学课程,本课程的建立将填补这类课程国内外的空白,为我国的化学教育及未来的化学研究发展提供助力。

(2) 自主构建化学信息数据库,推动化学研究从传统“查文献预测实验”模式到“大数据科学预测实验”模式的转变。

(3) 人工智能化学课程将整合现有的计算化学、虚拟仿真等教学资源,形成全新的信息化教学体系。

(4) 构建AI+化学化工虚拟仿真平台,为更多学者研究机器学习在涉化领域的应用提供学习平台。

## 参 考 文 献

- [1] 中国大学生就业, **2018**, 399 (9), 4.
- [2] 教育部. 中国信息技术教育, **2018**, 284 (8), 2.
- [3] 徐雷. 科学, **2017**, 69 (3), 1.
- [4] 姚琳, 石志国. 中国大学教学, **2019**, 350 (10), 19.
- [5] 叶民, 钱辉. 高等工程教育研究, **2017**, No. 4, 5.
- [6] 杜圣东, 杨燕. 计算机教育, **2020**, No. 7, 106.
- [7] 李德毅, 马楠. 高等工程教育研究, **2017**, 17 (5), 8.
- [8] Romero-Rivera, A.; Garcia-Borràs, M.; Osuna, S. *Chem. Commun.* **2016**, 53 (2), 284.
- [9] McKinney, S. M.; Sieniek, M.; Godbole, V.; Godwin, J.; Antropova, N.; Ashrafian, H.; Back, T.; Chesus, M.; Corrado, G. C.; Darzi, A; *et al.* *Nature* **2020**, 577, 89.
- [10] 朱博阳, 吴睿龙, 于曦. 化学学报, **2020**, 78 (12), 1366.
- [11] 刘献君. 高校教育管理, **2020**, 14 (1), 1.