附件目录

成果名称： 面向通用 AI 的语言智能教学体系建设与实践

成果完成人： 宗成庆，赵军，张家俊，刘康，何世柱，刘斌，陈玉博，赵阳，王少楠

成果完成单位： 中国科学院大学

一、成果总结

二、支撑材料目录

1 荣获奖励

1.1 教育教学奖

1.2 科研成果奖

1.3 最佳论文奖

2 教材封面

3 教材应用与课程推广证明

4 竞赛获奖

一、成果总结

1、成果背景

在 DeepSeek-R1、GPT-o3 等大模型持续突破的今天，人工智能正重构语言智能的技术范式。面对这一历史性变革，我国人工智能教育正迎来政策驱动与产业需求双重赋能的战略机遇期。教育部《高等学校人工智能创新行动计划》推动高校建设人工智能一级学科，加强基础研究、技术创新和人才培养。北京市推动“人工智能+”行动计划（2024-2025 年），实施“人工智能+教育”标杆工程，培育跨学科教育大模型平台，提供智能备课、课程分析等工具。语言智能作为人工智能皇冠上的明珠，既是全球科技竞争的主战场，也对高等教育提出了前所未有的挑战。

在上述背景下，当前语言智能技术发展凸显三大特点：1）技术演进呈现指数级加速趋势，大模型技术迭代快，学界和业界竞争激烈，导致创新周期和技术更新持续压缩。2）学科融合突破单一模态认知局限，语言学与神经科学、认知心理学、知识工程等领域深度交叉融合。3）实践难度大、要求高，大模型方法创新多源于工程实践总结和发现，实战能力成为语言智能技术发展的核心驱动力。

然而，传统语言智能人才培养体系与实际需求的错位现象日益凸显：1）知识架构存在传统范式固化的困境。一方面，部分高校仍沿用传统的教学体系，将大模型技术简化为选修模块，导致学生错失技术前沿；另一方面，部分激进革新者则完全抛弃传统方法范式，造成学生存在断层化知识体系，难以支撑持续创新。2）在学科交叉层面，语言智能本是语言学、认知科学、计算机科学、神经科学的交叉学科，但现实中往往呈现学科孤立现象，上述学科和课程的孤立现象阻碍了科学创新和跨学科发展。3）

实践训练存在能力鸿沟，现有课程实验多聚焦单任务、小规模数据集训练等局部环节，缺乏对数据清洗、模型训练、系统部署等全链路的实战锻炼。更值得关注的是，对前沿技术如大模型能力边界探索、对齐训练等关键领域，尚未形成体系化的实践培养方案，导致校企间存在显著的失衡。

中国科学院大学人工智能学院语言与知识处理教学团队经过多年探索构建了面向通用人工智能的语言智能培养体系，形成了“基础与前沿并举、跨学科融合交叉、研究与实践结合”的语言智能教学体系。该体系以适应新时代语言智能技术为核心目标，通过重构课程体系、创新交叉学科框架、搭建实践网络、撰写相关教材，为培养创新型语言智能人才提供坚实支撑。

2、成果内容和主要解决的教学问题

针对语言智能人才培养体系存在的知识滞后性、学科割裂性和实践脱节性问题，构建了“基础与前沿并举、跨学科融合交叉、渐进式实践强化”的语言智能教学体系。该体系具有多学科交叉、传统方法与前沿技术平滑衔接、理论方法学习与实践能力培养相结合特色的课程布局，从专业基础课程教学到实践，从核心专业课到专业选修课、研讨课教学，再从专业课程教学到研究生科研能力培养，全链条打通了自然语言、知识、情感以及神经语言学等相关知识点之间的壁垒，并与人工智能原理、模式识别与机器学习等相关课程协调，为语言智能人才培养持续赋能。

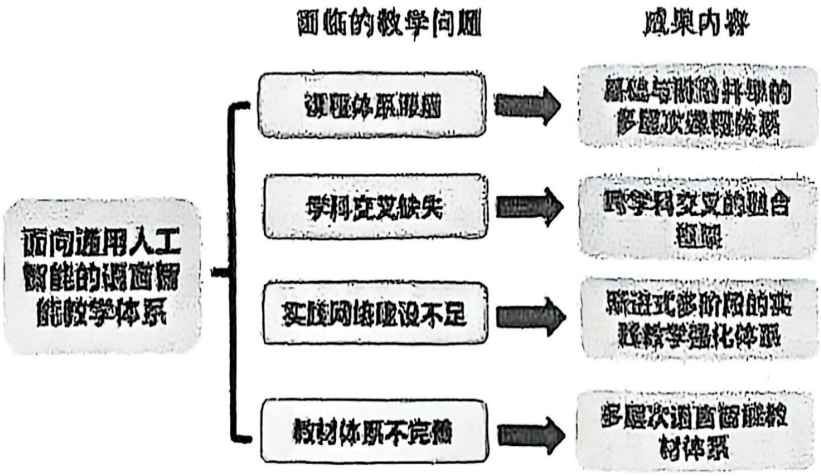

图 1 主要解决的教学问题和成果内容

(1)基础与前沿并举的多层次课程体系。该教学体系秉持“厚基础、强前沿”的理念，形成基础奠基、交叉进阶、前沿突破三层培养架构。在基础奠基阶段设置《语音信息处理》、《自然语言处理与大模型》、《信息检索》、《知识工程》等课程，通过项目制课程培养代码实现、系统构建及工程转化能力，实现学用贯通。既强调传统方法，也介绍前沿技术。交叉进阶阶段，设置《认知语言学》、《自然语言处理与应用》、《情感计算》等课程，着重培养跨学科融合思维。在前沿突破阶段，设置《大语言模型技术方法与应用》，实现更新研讨课程，着重培养学生创新能力。

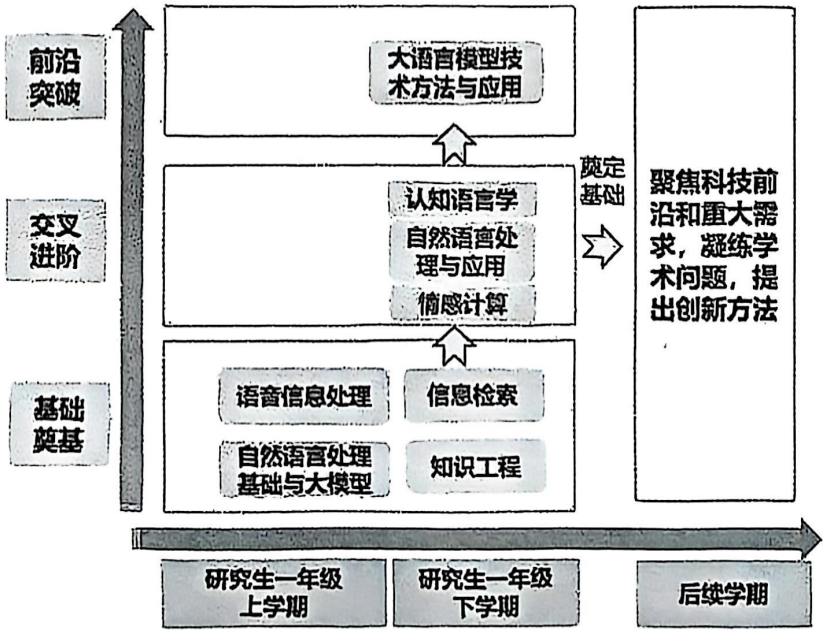

图 2 基础与前沿并举的多层次课程体系

(2)跨学科深度耦合的教学框架。为突破传统学科边界，创新性地构建语言技术、认知神经、知识工程、语音处理的多学科融合培养体系，形成从基础理论到前沿技术的跨学科培养链路。打破了将自然语言处理作为一门孤立课程进行教学的传统模式，全链条打通了语音语言处理、知识体系建设和情感情绪分析以及语言学习和理解的神经基础探索等相关知识点之间的壁垒，力求培养跨学科融合的复合型创新人才。在认知与神经科学课程让学生了解语言背后的认知和神经基础。文本处理技术相关课程培养学生形成多层面的语言处理知识体系。知识工程相关课程培养学生在通用 AI 时代的知识获取、表示和利用能力。在语音处理方向相关课程覆盖声学建模、情感语音合成等语音技术专题，最终培养跨学科复合型人才。

(3)渐进式多阶段的实践教学强化体系.构建课程作业、实践课程、实战锻炼的渐进式多阶段实践强化教学体系。理论课中提升了实践作业的比

重，巩固核心知识点。课程实践课程涵盖了对话系统、知识图谱构建、大语言模型等典型实践场景，聚焦全流程实践，培养学生在真实项目的实践能力；鼓励学生参与行业竞赛，通过实战锻炼让学生进行实战训练。与北京市超算中心共建语言智能算力基座，支持大模型参数模型训练的异构计算平台，使学生熟悉产业级训练架构。

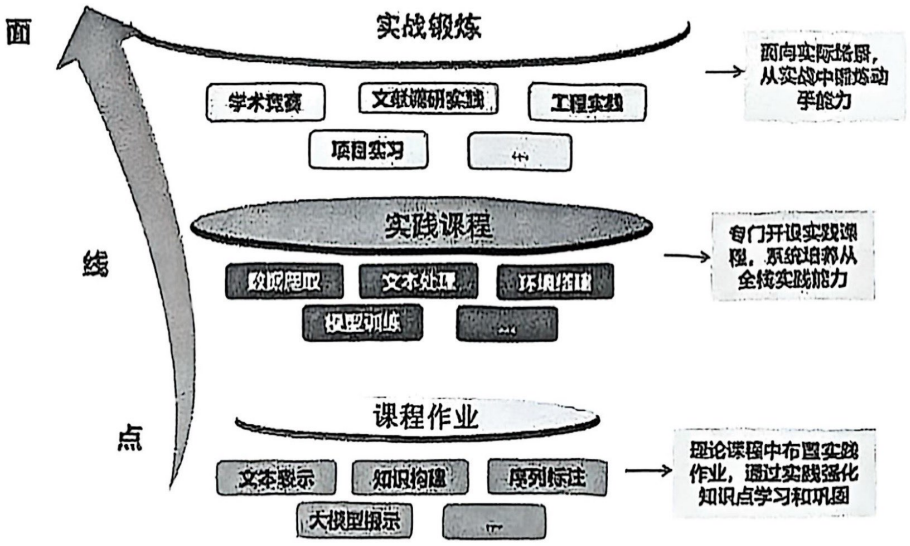

图 3 渐进式多阶段的实践教学强化体系

(4)撰写出版一系列自然语言、语音与知识处理的系列教材。围绕语言智能领域的课程体系建设目标，自 2008 年撰写出版了系列教材，包括《统计自然语言处理》、《文本数据挖掘》、《Text Data Mining》、《自然语言处理基础与大模型—案例与实践》、《知识图谱》、《知识图谱:算法与实践》、《情感计算理论与方法》。本系列教材和教学辅导书在国内产生了广泛的影响。例如，宗成庆老师独立撰写的《统计自然语言处理》第 1 版和第 2 版分别被 4 次和 17 次印刷发行，累计发行 3.5 万余册，成为本领域国内最早、影响力最大的关于统计自然语言处理方法的教学辅导书和研究参考文献。他讲授自然语言处理的教学视频在超星学术平台上被 34 万多注册用户和

1 万 3 千多课程平台点击播放。宗成庆老师被邀请参加了教育部 101 系列教材《自然语言处理》的编写工作。宗成庆作为第一作者、张家俊作为第三作者撰写的《Text Data Mining》于 2021 年由 Springer 出版社出版以来，已被点击下载 4.6 万余次，在本领域产生了很大的影响。赵军老师牵头撰写的《知识图谱》已被印刷 5 次，累计印数为 1 万 5 千余册，被国内很多知名高校选作教材或者参考书。宗成庆作为第一作者、赵阳作为第二作者撰写的《NLP 基础与大模型-案例与实践》，与百度合作建设面向自然语言处理的实践案例和平台，强化学生实践能力。

该体系主要针对以下教学难题:

(1)课程体系滞后于技术迭代问题。语言智能技术呈现指数级演进特征，而传统教学体系传统课程体系更新存在滞后性，导致教学内容与学术和产业前沿形成显著代际差，亟需构建基础理论与前沿技术的动态平衡机制。

(2)跨学科交叉融合缺失问题。语言智能已演变为语言学、认知科学、计算机科学的深度交叉领域，但现有课程存在明显“技术孤岛”现象，需打破学科壁垒，构建跨学科融合培养体系。

(3)实践网络缺位问题。语言智能作为强实践性学科，传统教学偏重理论传授，实践环节存在断点式缺陷。需构建全链路实践网络，有效弥合理论知识与实践应用的鸿沟。

3、成果采用的教学方法

针对语言智能人才培养体系存在的知识滞后性、学科割裂性和实践脱节性问题，构建了“基础与前沿并举、跨学科融合交叉、研究与实践结合”

的语言智能教学体系。

（1）基础与前沿并举的多层次课程体系。

在本科阶段课程设置中，构建“基础奠基-交叉进阶-前沿突破”三层培养架构具体优化方案如下，并针对本科与研究生差异化培养定位，实施分阶段能力跃升课程计划。研究生阶段，核心理论模块开设《自然语言处理基础与大模型》《知识工程》《语音信息处理》《情感计算》四门主干课程，系统构建算法模型、知识表示、声学处理、情感建模的理论框架；前沿研讨模块设置《认知语言学》《大语言模型技术方法与应用》《信息检索》等专题课程，通过文献研读、技术复现等形式，培养文献总结和分析能力；针对非全日制研究生开设《自然语言处理与应用》课程，聚焦大模型微调、领域适配、部署优化等，突出非全日制研究生学以致用的培养特色。

（2）跨学科深度耦合的教学框架。

为突破传统学科边界，创新性地构建“认知神经-语言技术-知识工程-语音处理”多学科融合培养体系，形成从基础理论到前沿技术的跨学科培养链路。在认知与神经科学领域，通过《认知语言学》、《语言与脑数据处理》课程让学生了解语言背后的认知和神经基础。文本处理模块设置《自然语言处理》、《自然语言处理实践》和《大语言模型技术方法与应用》课程。知识工程方面设置《知识工程》《信息检索》课程，培养学生在通用AI时代的知识获取、表示和利用能力。在语音处理方向，《语音信息处理》课程覆盖声学建模、端到端识别、情感语音合成等前沿专题，通过《情感计算》课程实现多模态联合建模。最终培养跨学科复合型人才。

（3）渐进式多阶段的实践教学强化体系。

构建课程作业、实践课程、实战锻炼的渐进式多阶段实践强化教学体

系。在理论教学中布置实践作业，要求学生在掌握具体知识点后，完成实践作业，强化知识点学习和巩固；专门开设《自然语言处理实践》课程，系统培养从数据爬取、清洗标注、模型优化到部署推理的全栈开发能力；与北京市超算中心共建算力基座，部署使学生直接接触产业级训练平台，锻炼学生在项目层面的一体化训练。竞赛培养采用双轨制：课内设置《文献复现与创新实践》课程，实施文献调研、代码复现、实验分析的科研闭环训练；课外鼓励学生参与国内外权威竞赛。

4、创新点

该教学体系创新呈现三大特色：

（1）基础-前沿双螺旋递进式课程体系。

构建“理论筑基-交叉进阶-前沿突破”三层架构，形成从基础到前沿的完整培养链路。设置《大语言模型技术方法与应用》《认知语言学》等前沿课程，追踪通用 AI 与多模态技术演进。非全日制研究生特设《自然语言处理与应用》工程课程，突出技术转化能力培养。

（2）跨学科深度耦合教学框架。

打破传统学科界限，构建“认知神经-语言技术-知识工程-语音处理”融合体系。通过《语言与脑数据处理》课程联结语言学与神经科学，《大语言模型技术方法与应用》融合语言技术与计算机前沿，《信息检索》与《情感计算》课程实现知识工程与语音处理的交叉创新，培养兼具语言认知素养与 AI 开发能力的复合型人才。

（3）渐进式多阶段实践强化生态。

建立课程作业、实践课程、实战锻炼三阶实践体系：理论教学侧重实践作业，强化知识点学习和巩固。《自然语言处理实践》课程培养全方位

开发能力；通过《文献复现与创新实践》课程实施科研闭环训练，并鼓励参与机器翻译、情感计算等国际赛事。

5、成果推广应用

本成果在教育教学实践、教材体系构建及创新人才培养领域取得系统性突破，形成多维度育人效能并产生广泛示范效应。

团队教育建设与改革成效显著，取得了多项教育教学奖励。宗成庆研究员获得中国科学院大学李佩教学名师、中国科学院大学“领雁奖章-振翅奖”、宝钢优秀教师奖、中国科学院优秀导师奖、中国科学院教育教学成果奖二等奖、北京市优秀教师等若干荣誉称号。宗成庆、赵军、张家俊、刘斌 4 名团队教师获得中国科学院朱李月华优秀教师奖。《自然语言处理》、《知识图谱》、《情感计算》、《自然语言处理及应用》等课程 4 次荣获中国科学院大学校级优秀研究生课程。

教材建设成果丰硕，形成多层次语言智能教材体系，产生广泛应用和推广。团队共主编教材 7 本，包括《统计自然语言处理》、《文本数据挖掘》、《Text Data Mining》、《自然语言处理基础与大模型—案例与实践》、《知识图谱》、《知识图谱:算法与实践》、《情感计算理论与方法》。教材体系呈现三大特色：1）基础理论与前沿技术融合；2）中英双语教材配套，满足国际化教学需求；3）产教融合案例丰富。相关教材被国内百余所高校采纳。例如《统计自然语言处理》第 1 版和第 2 版分别被 4 次和 17 次印刷发行，累计发行 3.5 万余册，成为本领域国内最早、影响力最大的关于统计自然语言处理方法的教学辅导书和研究参考文献。《知识图谱》该书目前已印刷 5 次，累计印数为 15054 册，被中国科学院大学、北京师范大学等国内十几所高校选作教材或者参考书。

创新人才培养成效显著。研究团队所培养学生入选国家自然科学基金优青项目 3 人，北京智源青年科学家 3 人，中国科协青年人才托举工程 3 人，中科院青年创新促进会 4 人，获中国中文信息学会优秀博士论文 3 篇。研究成果荣获多项国内外著名会议的最佳论文，例如获 COLING2014 最佳论文(CCF B 类)、ISWC2023 最佳张贴论文(CCF B 类)及 3 篇 CCKS、3 篇 CCL 国内顶级会议最佳论文，CCMT 最佳论文。学生竞赛能力持续领跑，在自然语言处理、医疗文本处理、知识图谱、语音处理等领域的竞赛屡获突破，累计或得国内外评测冠军 14 项。

二、支撑材料目录

1、 荣获奖励 1

1.1 教育教学奖 1

1.1.1 北京市优秀教师（宗成庆） 1

1.1.2 宝钢优秀教师奖（宗成庆） 2

1.1.3 国科大领雁奖章（宗成庆） 3

1.1.4 中科院教育教学成果奖二等奖（宗成庆） 4

1.1.5 2018 中科院优秀导师奖（宗成庆） 5

1.1.6 2019 中科院优秀导师奖（宗成庆） 6

1.1.7 中国科学院大学李佩教学名师（宗成庆） 7

1.1.8 国科大-华为智能基座优秀教师（刘斌） 8

1.1.9 朱李月华优秀教师 2010（宗成庆） 9

1.1.10 朱李月华优秀教师 2020（赵军） 10

1.1.11 朱李月华优秀教师 2022（张家俊） 11

1.1.12 朱李月华优秀教师 2024（刘斌） 12

1.1.13《自然语言处理与应用》2023 年校级优秀课程（赵阳） 13

1.1.14《知识图谱》2021 年校级优秀课程（赵军） 14

1.1.15《自然语言处理》2019 校级优秀课程（宗成庆） 15

1.1.16《自然语言处理与应用》2023 年院级优秀课程（赵阳） 16

1.1.17《知识图谱》2021 年院级优秀课程（赵军） 17

1.1.18《知识图谱与语义计算》院级优秀课程（赵军） 18

1.1.19《情感计算》院级优秀课程（刘斌） 19

1.2 科研成果奖 20

1.2.1 国家科技进步奖二等奖（宗成庆） 20

1.2.2 中国专利优秀奖（刘斌） 21

1.2.3 中国专利优秀奖（宗成庆） 22

1.2.4 教育部自然科学奖二等奖（宗成庆） 23

1.2.5 北京市科学技术奖一等奖 2021（宗成庆、张家俊） 24

1.2.6 北京市科学技术奖一等奖 2019（赵军、刘康、何世柱、陈玉博） 26

1.2.7 北京市科学技术进步奖二等奖 2014（刘斌） 30
1.2.8 自治区科学技术进步奖二等奖 2019（宗成庆） 31
1.2.9 中国电子学会技术科技进步一等奖（宗成庆） 33
1.2.10 中国电子学会技术发明一等奖（刘斌） 34
1.2.11 国务院特殊津贴（宗成庆） 35
1.2.12 国务院特殊津贴（赵军） 36
1.2.13 中国中文信息学会钱伟长中文信息处理科学技术奖 2014（宗成庆） 37
1.2.14 中国中文信息学会钱伟长中文信息处理科学技术奖 2018（赵军、刘康、何世柱、陈玉博） 38
1.2.15 中国中文信息学会钱伟长中文信息处理科学技术奖 2020（何世柱） 40
1.2.16 IEEE Fellow（宗成庆） 41
1.2.17 ACL Fellow（宗成庆） 42
1.2.18 青年人才托举工程（陈玉博） 43
1.3 最佳论文奖 44
1.3.1 第十九届全国机器翻译大会 44
1.3.2 第 9 届国际计算与模式识别会议最佳论文 45
1.3.3 第 16 届全国人机语音通信会议最佳论文 46
1.3.4 COLING2014 best paper 47
1.3.5 ISWC2023-获奖证书 48
1.3.6 CCKS 2017 最佳论文 49
1.3.7 CCKS 2019 最佳论文 50
1.3.8 CCKS 2020 英文最佳论文 51
1.3.9 NLP-NABD2016 最佳论文 52
1.3.10 CCL2019 最佳论文 53
1.3.11 CCL2020 最佳论文 54
2、 教材封面 55
2.1《统计自然语言处理》 55
2.2《文本数据挖掘》 56
2.3《文本数据挖掘》（第 2 版） 57

2.4《Text_Data_Mining》 58
2.5《自然语言处理基础与大模型——案例与实践》 59
2.6《知识图谱》 60
2.7《知识图谱：算法与实践》 61
2.8《情感计算理论与方法》 62
2.9《情感分析-挖掘观点、情感和情绪》 63
3、 教材应用与课程推广证明 64
3.1《统计自然语言处理》教材使用情况 64
3.2《文本数据挖掘》教材使用情况 65
3.3《文本数据挖掘》（第 2 版）教材使用情况 66
3.4《自然语言处理基础与大模型-案例与实践》教材使用情况 67
3.5 超星学术视频点击播放次数证明（2025.04） 68
3.6《知识图谱》出版证明 69
3.7《知识图谱：算法与实践》出版证明、版权页、目录、文献页、内容页 70
3.8 《情感计算理论与方法》教材使用情况 75
3.9《情感分析-挖掘观点、情感和情绪》出版说明版权页、文献页、内容页 76
4、 竞赛获奖 79
4.1 第 1 届多模态情感分析挑战赛冠军 79
4.2 第 2 届多模态情感分析挑战赛冠军 80
4.3 第 3 届多模态情感分析挑战赛冠军 81
4.4 第 5 届多模态情感分析挑战赛冠军 82
4.5 第 6 届国家微表情挑战赛冠军 83
4.6 KDD CUP2011 84
4.7 CCKS 2020 评测技术创新奖 85
4.8 CCKS 2020 评测一等奖 86
4.9 万创杯证书 87
4.10 AAAI2021 “缩写词抽取与消歧”评测冠军 88
4.11 CCKS 2021 Aminer 学者画像-获奖证书 89
4.12 CCKS 2021 “医疗科普知识答非所问识别”赛道评测任务冠军 90

4.13 “未来杯”人工智能与机器人项目展示活动（高校组） 91
4.14 “2023_智研”杯大模型国防应用邀请赛 应用潜能展示优胜奖 92