

郝爽



副教授、硕士生导师

个人主页: <https://faculty.bjtu.edu.cn/9279/>

联系方式: haoshuang@bjtu.edu.cn

教育及工作背景

- 2022.1-至今 北京交通大学计算机学院副教授
- 2013.9-2018.7 清华大学计算机系直博

研究方向

- 面向人工智能的数据管理
- 新型数据库技术
- 大模型赋能的数据管理

研究成果

- 在国际顶级会议VLDB、ICDE、AAAI等和国际顶级期刊TKDE、VLDBJ等发表高水平论文二十余篇
- 承担包括国家重点研发计划、国家自然科学基金等国家级项目五项
- 相关工作获得麻省理工大学、斯坦福大学、普度大学、爱丁堡大学、清华大学、北京大学等国内外顶尖学府同领域学者的正面引用

社会兼职

- 中国计算机学会数据库专委会执行委员
- 数据库与数据挖掘领域国际顶级会议VLDB、ICDE、KDD、AAAI、IJCAI程序委员会委员

课题组介绍

- 课题组隶属于北京交通大学计算机学院计算机科学与工程系、交通数据分析与挖掘北京市重点实验室
- 成员年轻, 成果丰富, 积极向上, 自由平等
- 中等规模, 每位研究生均亲自指导, 鼓励冲击CCF-A类会议和期刊
- 每周组会+周报, 周六日、寒暑假期间不安排团队活动 (看重是否完成既定目标, 而非在实验室积累时间)
- 科研经费充足, 硬件环境支撑到位, 每月补贴高于学院平均水平并有额外奖励机制, 论文录用后支持参加国内外高水平学术会议
- 针对硕士毕业后就业的学生, 支持研二暑假外出实习
- 每学年大规模实验室团建活动一次

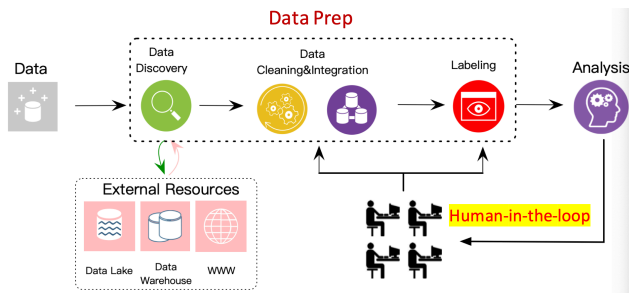


期望中的你

- 有好好做事的决心与认真做事的态度, 有志参与高水平研究和项目、收获高水平成果, 拒绝摆烂和躺平
- 有科研经历或者论文投稿经历、ACM获奖经历或者项目实践经历
- 扎实的算法基础, 优秀的编程能力, 较好的英语水平
- 良好的沟通能力和团队协作能力

毕业后的你

- 具有独立解决问题的信心和能力, 准确知晓前沿技术方向
- 优秀的编程能力和项目实习经历→收获一份心仪的工作
- 优秀的科研素质和研究成果 (以第一作者身份发表CCF-A类论文) →收获去其它高校深造的机会 (团队同时提供转博机会)
- 拥有一群志同道合的小伙伴

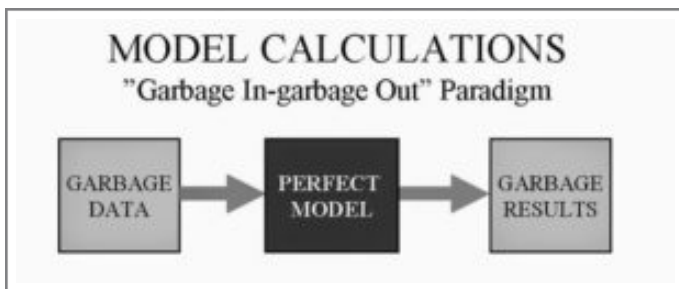


面向人工智能的数据管理

- 人工智能模型从数据中学习，现阶段人工智能是**数据智能**，没有数据就没有智能
- 人工智能技术从“以模型为中心”发展为“以**数据为中心**”
- 当AI系统从**实验室**走向**真实应用**，**数据质量**的好坏起到非常重要的作用，数据质量已成为人工智能模型在落地过程中的一大瓶颈

$$\text{AI System} = \text{Data} + \text{Code}$$

- 课题组的研究聚焦在**数据**上，突破**数据质量低**、**数据准备成本高**的问题，深入研究人工智能技术走向真实产业应用的关键因素
- 围绕着训练**数据发现**、**清洗与集成**、**数据标注**、**模型容错训练**等诸多科学问题展开研究，涵盖**结构化数据**、**非结构化数据**、**多模态数据**



新型数据库技术

数据库是数字化时代的核心基础设施，被称为信息技术系统皇冠上的明珠。目前数据库人才缺失，数据库国产化任重而道远。课题组紧跟数据库领域发展趋势，聚焦如下研究方向：

• 人工智能赋能的数据库技术

利用人工智能技术使数据库更加智能化，增强数据库功能，同时简化用户操作

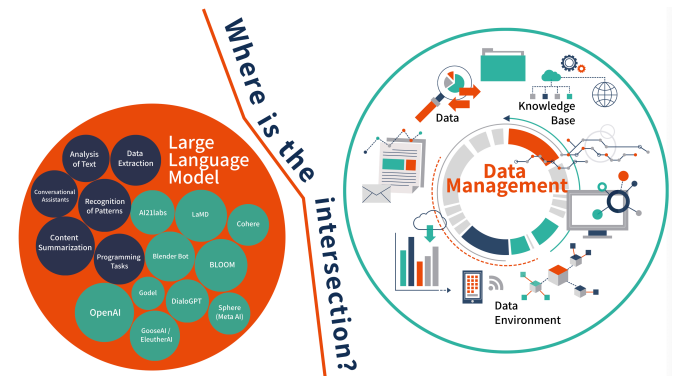
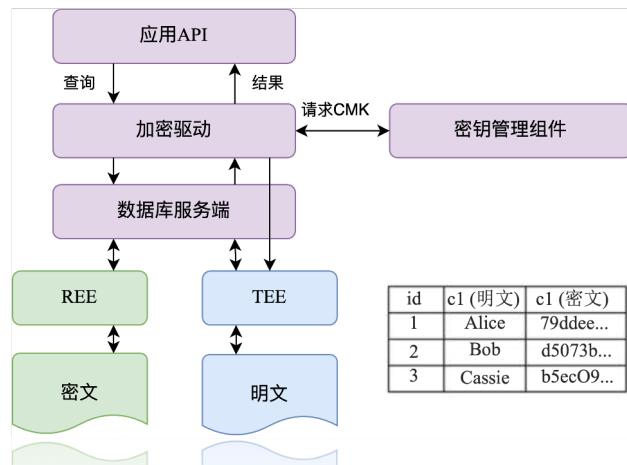
• 向量数据库

研究向量数据存储和向量检索算法，使得向量数据库成为大模型的长期记忆体



• 软硬融合的全密态数据库

融合软件加密算法和可信硬件的优点，保证数据库机密性的同时，高效支持SQL功能，推动全密态数据库的技术革新



大模型赋能的数据管理

- 大语言模型是目前人工智能领域最热门的一个话题，大模型从大量标记和未标记的数据中捕获知识，以实现对各种任务进行高效的处理
- 大模型已经在多个领域都有广泛应用，成为赋能企业效率提升的关键驱动力
- 当前**大模型和数据库融合**的趋势越来越明显，大模型与数据库是人工智能时代的**双向助推力**，数据库与大模型结合势必成为未来二者发展的一大趋势
- 本课题主要研究如何构建大模型赋能的数据管理的技术体系，**包括但不限于**大模型赋能的数据治理、数据分析、NL2SQL、查询重写、数据库系统诊断等

