

陈沁宇

19959885958 | chenqinyu@pku.edu.cn
Wechat: QinyuChen_99 | https://qinyuchen.site



教育经历

北京大学 - 计算机科学与技术 硕士 计算机学院-计算语言所 & 人工智能创新中心 2022年09月 - 2025年07月
北京大学 - 计算机科学与技术 本科 信息科学技术学院 2017年09月 - 2022年07月

实习经历

微软 - 算法实习生 Bing Ads Selection 2023年08月 - 2024年02月

- 结合大语言模型和推荐系统算法改进Bing广告召回模型，独立完成全流程数据处理、模型优化、模型训练、效果评估。
- 在离线测评中有效提升广告选择结果的相关性、召回率和准确率。新增重排模块能在较小额外开销前提下，相比原先提升36%召回率和40%相关性得分。

字节跳动 - 算法工程师 抖音搜索 2021年12月 - 2022年09月

- 利用深度推荐算法，优化抖音搜索用户行为模型。独立负责多个双月项目，包括引入多模态信息优化搜索结果、对用户点击行为进行概率建模、采用知识蒸馏利用Session级行为特征、引入新的行为目标等；
- 在AB测试中显著提升点击率、完播率和留存；模型成功部署上线，投入使用。

深度求索 - AGI研发 2024年03月 - 至今

- 参与开发Agent框架、代码领域专用模型
- LLM预训练相关研究

项目经历

Kaggle Optiver - Trading at the Close 股价预测比赛 银牌 - 算法研发 2023年10月 - 2023年12月

- 处理股票收盘阶段的交易数据，探索时序特征、截面特征的使用，调研高频特征与低频特征相结合
- 负责模型参数调优、超参搜索和推理加速。

第四届最高法“法研杯”法律人工智能挑战赛 冠军 - 算法研发 2021年10月 - 2022年01月

- 根据法律文本特点，使用针对性的文本增强方法，采用端到端的方式进行特征提取与刑期预测；
- 引入RoPE增强Transformers，并使用真实刑期分布作为后验修正。获得刑期预测赛道冠军。

移动云应用创新开发大赛暨“睿聚杯”亚军 - 算法研发 2021年08月 - 2021年10月

- 采用了针对法律文本的预处理、无监督预训练、标签平滑、层次处理、模型融合等技术；
- 参与民事案由多层级分类赛道，初赛第2，复赛第1，总榜亚军。

论文发表

Domain Adaptation via Maximizing Surrogate Mutual Information IJCAI2022-Long Pre
Exploring In-Context Learning for Knowledge Grounded Dialog Generation EMNLP2023 Findings
Selecting Large Language Model to Fine-tune via Rectified Scaling Law ICML 2024 Poster
Retrieval-based Full-length Wikipedia Generation for Emergent Events

学工经历

北京大学研究生院学位办 - 学生助理 2023年02月 - 至今

协助汇总、审核全校博导硕导招生权限，协调分析学科评估结果，主要负责数据分析和可视化呈现。

基于深度学习的自然语言处理 - 课程助教 2022年09月 - 2024年01月

2022年、2023年连续担任北大元培学院通用人工智能班必修课助教，负责课程内容设计和课程项目考核。

北京大学吉他协会 - 会长 2020年01月 - 2020年12月

协调宣传、迎新活动，平时组织教学班，学期末举办原声专场和乐队专场。当年新增会员人数、孵化乐队数为近年最多。

技能奖项

- 编程技能: Python / Pytorch / Tensorflow / C++ / Spark / Sql / Office
- 语言考试: TOFEL 110 (R30 L28 S25 W27), CET-6: 632, CET-4: 659
- 奖项荣誉: 深交所奖学金, 研究生学业奖学金, 校三好生, 优秀科研奖
- 兴趣爱好: 吉他、篮球、羽毛球、健身、电影、咖啡