

姓名: 沙九

甘肃甘南 | 中共党员 | 1994 年 9 月 | 2021 年 7 月硕士毕业| ☎: (+86)18810979033(微信同步)

✉: 18810979033@163.com 主要研究方向: 自然语言处理 (NLP)、机器学习(ML)

期望岗位: NLP/CV/机器学习/深度学习/软件开发/数据分析/



## IT 技能

- 熟悉:Python/PyTorch/Linux/Java; 熟练掌握: Transformer/TensorFlow/Oracle/C/C++;其他深度/机器学习算法

## 教育背景

- 北京理工大学	计算机科学与技术专业(平均成绩: 92.64/100.00)	硕士	2018.09-2021.07
- 中央民族大学	计算机科学与技术专业(平均成绩: 95.52/100.00)	学士	2014.09-2018.07

## 项目/科研经历

- 多策略特定领域的智能机器翻译系统 2019.09-2020.01 项目
  - 项目描述: 为某军区通信监测研发一套特定领域内多策略融合方式的机器翻译系统。
  - 项目职责: 担任组长、项目落地模型的选择, 算法的设计, 系统的搭建、程序的实现。
  - 项目业绩: OpenNMT 框架/基于 Transformer 模型/用 Flask—Web 搭建翻译平台, 支持有选择性的二次翻译。
- 面向法言法语的稀缺资源机器翻译技术研究及设备研发 2018.10-2021.07 项目
  - 项目描述: 研究面向法言法语的多语种机器翻译技术与互译便携式设备。
  - 项目职责: 多维度丰富单语, 利用句法特征分析语言建模、信息抽取以及特征工程等聚类 and 泛化司法领域数据。
  - 项目业绩: 集成多语种的便携式终端, 能够支持 6 个语种 12 个方向的翻译系统。
- 融合大数据与人类常识的开放域多语言知识图谱构建 2019.03-2020.11 项目
  - 项目描述: 大数据驱动的自然语言理解、问答和翻译。
  - 项目职责: 研究大数据驱动的世界/常识知识获取与融合方法, 建立大规模、高质量、融合常识知识的多语言知识图谱。
  - 项目业绩: 为大数据驱动的自然语言理解、问答和翻译提供基础知识资源。

## 实习/工作经历

- 公司名称: 北京机械工业自动化研究所软件事业部 2021.07-至今 工作
  - 职位名称: 技术研发工程师
  - 工作描述: 负责微服务后端的研发业务
  - 工作职责: 参与设计后端研发业务, 通过 Springboot Mybatis 实现数据(Oracle、Mysql)交互为前端提供服务。
  - 完成业绩: 完成两个后端研发设计; 主导数据字典、计量单位维护、动态链接数据源等后端研发; 撰写后端研发标准规范并提供培训。
- 公司名称: 华为技术有限公司 2020.06-2020.10 实习
  - 职位名称: Web 异常检测 AI 算法实习生。
  - 工作描述: 研究和挖掘注入类的 web 攻击, 自动化恶意流量家族提取, 利用 AI 识别 Web 的攻击检测。
  - 工作职责: 结合自然语言处理, 通过机器学习深度学习等。文本分类、情感分析、舆情监测以及信息检索等。
- 公司名称: 北京智源人工智能研究院 2020.01-2020.05 实习
  - 职位名称: AI 算法实习生。
  - 工作描述: 挖掘学者的研究兴趣。
  - 工作职责: 基于海量的学者数据信息分析, 构建从多维度获取特征并能够自动生成每位学者的研究兴趣。
  - 完成业绩: 通过已提供的学者信息特征及外部嵌入描述, 构建一套能够自动生成对应学者研究爱好的系统。

## 竞赛/获奖/论文情况

- 十一届中国计算机博弈锦标赛冠军;CWMT2018/CCMT2019 藏汉翻译第二名/第四名;“第十三届中国大学生年

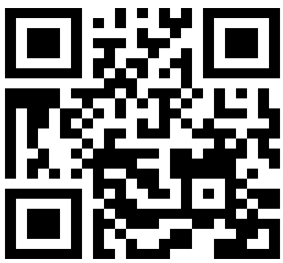


度人物” 候选；北京市三好学生；宝钢优秀学生奖；国奖；专一；优秀学生干部等，北京市优秀毕业生，优秀硕士生。

- 2022 年 ACM 《Integrating Pre-training model into NMT with Bi-Directional Feature Transformation》。
- 2021 年 普通发明专利《一种利用半自回归融合领域术语的低资源机器翻译方法》。
- 2020 年 CCL 《面向司法领域的高质量开源藏汉平行语料库构建》并获最佳论文奖。
- 2020 年 《Revisiting Back-Translation for Low-Resource Machine Translation Between Chinese and Vietnamese》 发表于 IEEE。
- 2019 年 《多策略切分粒度的藏汉双向神经机器翻译研究》 被 CCMT2019 会议录用并发表在《厦门大学学报(自然科学版)》。
- 2016 年 《藏文自动分词与词性标注研究》

### 自我评价

- 本人性格开朗，为人诚恳，乐于沟通，有团队精神。喜欢弹唱、足球。热爱钻研，心态良好，能积极面对工作中的困难。始终秉承“学以致用”的准则，学习能力强。希望能够在实践中持续学习，发挥自己的主动性、创造性，为公司的发展竭尽全力。

### 个人站点

		
博客	Github	微信公众号