

米尔阿迪力江·麦麦提

✉ miradel_51@hotmail.com 🌐 <http://www.miradeljan.com>

👤 政治面貌: 党员 🏠 籍贯: 新疆喀什市

🎓 教育经历

预科	2007.09 – 2008.07
语言学院 新疆大学	新疆乌鲁木齐
本科	2008.09 – 2012.07
计算科学与技术 信息科学与工程学院 新疆大学	新疆乌鲁木齐
硕士	2012.09 – 2015.06
软件工程 信息科学与工程学院 新疆大学	新疆乌鲁木齐
博士	2015.09 – 2021.06
计算科学与技术 计算机科学与技术系 清华大学	北京

💼 工作经历

中国科学院新疆理化技术研究所	2023.04 – 至今
研究助理	多语种信息技术研究室

- 地点: 新疆乌鲁木齐
- 项目: 低资源场景下的自然语言生成新的前沿技术
- 贡献: 我在跨语言小语种知识检索（语义检索）模型中引入了code-switching的方法，通过将正则化项整合到多语言预训练模型中，提高了跨语言语义检索模型的性能。同时，我尝试以完全无监督的方式利用上述提出的跨语言语义检索方法进行数据增强，而得到高质量伪数据，提出的算法模型能够带来非常显著地下降了增强后数据的语法错误（句法和语义错误）。除此之外，我还探索了一些在资源有限的情况下对大规模语言模型（LLM）进行高效监督微调（SFT）的方法。

阿里巴巴达摩院	2021.07 – 2023.03
NLP高级算法工程师	机器智能-智能对话与服务算法团队

- 地点: 浙江杭州
- 项目: DARAZ, LAZADA 以及 AE 多语言处理（跨境电商智能客服业务）
- 贡献: 主要负责DARAZ整体业务（算法owner）以及 AliExpress（商家）智能客服。在海外智能客服业务中，AliExpress 商家板块增加新的语言、中亚地区海外业务LAZADA意图分类模型取得了线下4个多点的提升以及南亚地区海外业务DARAZ的FAQ模型和意图分类模型取得了线下13个点和线上平均9个点的提升。

阿里巴巴达摩院	2020.07 – 2021.7
研究实习	机器智能-小蜜-多语言处理组

- 地点: 浙江杭州
- 项目: 乌尔都语-孟加拉语罗马化（跨境电商智能客服业务）
- 贡献: 实习期间提前完成主管安排的任务。提出了跟语言无关的基于混合策略的罗马化策略并解决了在DARAZ 等面向南亚地区的国际化电商平台上的核心问题。算法模型上线之后，在实际业务落地方面取得了明显的提升，主要表现在线上模型解决率的提升。

北京丝路合力投资管理集团	2019.07 – 2019.12
CTO	网络科技部

- 地点: 北京
- 项目: 多语言电商平台研发
- 贡献: 主要是负责以“一带一路”为中心的跨语言跨语境电商平台的研发、其中核心框架的设计并提供技术指导。此外，负责在线教育以及自媒体相关的平台开发管理并设计。

📁 项目经历

- 目前主持的项目列表
 - 国家自然科学**青年基金**项目2025.01 – 2027.12
 - “基于大模型的资源匮乏语言机器翻译方法研究”
 - * 级别：国家级
 - * 单位：中科院新疆理化所
 - * 经费：30万
 - * 状态：尚未开始
 - 自治区天池英才引进计划项目 (**青年博士计划**)2024.01 – 2027.01
 - “基于大规模语言模型的低资源神经机器翻译方法研究”
 - * 级别：省部级
 - * 单位：中科院新疆理化所
 - * 经费：75万 (制定的最高金额90万)
 - * 状态：在研中
 - 中国科学院**特别研究助理**资助项目2024.01 – 2026.01
 - “基于有监督微调的大模型低资源语言翻译性能优化”
 - * 级别：中科(院)级
 - * 单位：中科院新疆理化所
 - * 经费：60万 (制定的最高金额60万)
 - * 状态：在研中
- 主要参与的代表性项目列表
 - 清华大学
 - * 国家自然科学基金项目2014.01 – 2018.08
 - “面向三元空间的互联网中文信息处理理论与方法”
 - 级别：国家级 (973计划)
 - 导师：孙茂松院士
 - 经费：2438万
 - 状态：已结题
 - 贡献：主要负责开发神经网络维-汉双向机器翻译系统，已扩展到“一带一路”沿线国家多种语言之间的翻译服务。
 - * 国家自然科学基金项目2015.01 – 2017.12
 - “语言问题求解和答案生成关键技术以及系统”
 - 级别：国家级 (863计划)
 - 导师：刘洋教授
 - 经费：718万
 - 状态：已结题
 - 贡献：主要负责数据预准备、数据预处理以及相关模型训练，同时负责系统前端开发。
 - * 国家自然科学基金 (国际合作项目)2018.01 – 2020.12
 - “基于深度学习的汉葡机器翻译方法研究”
 - 级别：国家级
 - 导师：刘洋教授
 - 经费：200万
 - 状态：已结题

· 贡献：主要负责开发低资源神经机器翻译系统。

* **国家自然科学基金项目** 2015.01 – 2019.12
“面向多层次篇章语义的机器翻译方法研究与实现”

· 级别：国家级

· 导师：刘洋教授

· 经费：350万

· 状态：已结题

· 贡献：主要负责数据预准备、数据预处理以及相关模型训练。

* **国家自然科学基金 (重点) 项目** 2014.01 – 2018.12
“跨语言社会舆情分析基础理论与关键技术研究”

· 级别：国家级

· 导师：刘洋教授

· 经费：285万

· 状态：已结题

· 贡献：主要负责开发跨语言信息检索系统，分别爬虫并采集维吾尔语、藏语、蒙语网站，以及调用自主研发的小语种机器翻译接口而实现跨语言搜索。

* **国家自然科学基金项目** 2014.01 – 2016.12
“少数民族网络舆情综合分析与服务关键技术研究及应用示范”

· 级别：国家级 (科技部)

· 导师：刘洋教授

· 经费：78.4万

· 状态：已结题

· 贡献：主要负责训练少数民族语言文本情感分类模型以及优化。

* **高校-企业合作项目** 2017.01 – 2018.12
“基于深度学习的维汉机器翻译”

· 级别：企业级

· 导师：刘洋教授

· 状态：已结题

· 贡献：主要负责维-汉双向机器翻译引擎研发技术顾问，既是技术指导以及提供相应的code。

– 新疆大学

* **工业与信息化部科技项目** 2014.01 – 2017.01
“基于Android的维吾尔语语音控制系统”

· 级别：国家级

· 导师：吾守尔·斯拉木院士

· 状态：已结题

· 贡献：开发整套Android系统以及构建了维吾尔语槽语法（用于语音识别模块）文件。

* **国家自然科学基金项目** 2013.01 – 2017.01
“基于网络的维吾尔文舆情分析大型软件研究”

· 级别：国家级

· 导师：吾守尔·斯拉木院士

· 状态：已结题

· 贡献：开发维吾尔语词干提取预处理工具、人工词性标注平台、词性标注审核平台以及自动词干提取与词性标注系统。

* **国家自然科学基金项目** 2012.01 – 2015.01
“面向中西亚的出口软件开发”

- 级别：国家级
- 导师：吾守尔·斯拉木院士
- 状态：已结题
- 贡献：Win7 的阿拉伯语版校对统计及整理工作。

* 国家自然科学基金项目

2011.01 – 2014.01

“维吾尔文笔记鉴别研究”

- 级别：国家级
- 导师：吾守尔·斯拉木院士
- 状态：已结题
- 贡献：C++开发和系统测试。

🏆 学术成果

• 代表性论文

1. Mieradilijiang Maimaiti, Yang Liu*, Yuanhang Zheng, Gang Chen, Kaiyu Huang, Ji Zhang, Huanbo Luan and Maosong Sun, “Segment, Mask, and Predict: Augmenting Chinese Word Segmentation with Self-Supervision”, In International Conference on Empirical Methods in Natural Language Processing (EMNLP), p2068-2077, 2021 (CCF B类)
2. Mieradilijiang Maimaiti, Yang Liu*, Huanbo Luan, and Maosong Sun, “Data Augmentation for Low-Resource Languages NMT Guided by Constrained Sampling”, In International Journal of Intelligent Systems (IJIS), volume 37, p30-51, 2021 (SCI 2区 Top)
3. Mieradilijiang Maimaiti, Yang Liu*, Huanbo Luan, Zegao Pan, and Maosong Sun, “Improving the Data Augmentation for Low-Resource NMT Guided by POS-Tagging and Paraphrase Embedding”, In ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), volume 20, issue 6, p1-21, 2021 (CCF C类)
4. Mieradilijiang Maimaiti, Yang Liu*, Huanbo Luan, and Maosong Sun, “Enriching the Transfer Learning with Pre-trained Lexicon Embedding for Low-Resource Neural Machine Translation”, In Tsinghua Science & Technology (TST), volume 27, issue 1, p150-163, 2020 (SCI 1区)
5. Mieradilijiang Maimaiti, Yang Liu*, Huanbo Luan, and Maosong Sun, “Multi-Round Transfer Learning for Low-Resource NMT Using Multiple High-Resource Languages”, In ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), volume 18, issue 4, p1-26, 2019 (CCF C类)
6. Mieradilijiang Maimaiti, Shunpeng Zou*, Xiaoqun Wang, Xiaohui Zou, “How to Understand: Three Types of Bilingual Information Processing?”, In International Conference on Cognitive Systems and Signal Processing (ICCSSP), volume 1006, p3-16, 2019 (EI 检索)
7. Mieradilijiang Maimaiti, Xiaohui Zou*, “Discussion on Bilingual Cognition in International Exchange Activities”, In International Conference on Intelligence Science (ICIS), volume 539, p167-177, 2018 (EI 检索)
8. Mieradilijiang Maimaiti*, Yuanhang Zheng*, Ji Zhang, Fei Huang, Yue Zhang, Wenpei Luo, Kaiyu Huang, “Improving Multi-lingual Representation for Semantic Retrieval with Code-switching”, In International Journal of XXXXX, 2024 (SCI 1区), 在审
9. Zhe Li*, Mieradilijiang Maimaiti*, Jiabao Sheng, Zunwang Ke, Wushour Slamu, Qinyong Wang, Xiuhong Li, “An Empirical Study on Deep Neural Network Models for Chinese Dialogue Generation”, In Symmetry-Basel (Symmetry), volume 12, issue11, p1756-1772, 2020 (SCI 3区), 共同一作
10. Wenshen Xu*, Mieradilijiang Maimaiti*, Yuanhang Zheng, Xin Tang, and Ji Zhang†, “Auto-MLM: Improved Contrastive Learning for Self-supervised Multi-lingual Knowledge Retrieval”, In International Journal of YYYYYY, 2024 (SCI 1区), 共同一作, 在审
11. Jianhai Zhang, Mieradilijiang Maimaiti, Xing Gao, Yuanhang Zheng, and Ji Zhang*, “MGIMN: Multi-Grained Interactive Matching Network for Few-shot Text Classification”, In International Conference on

North American Chapter of the Association for Computational Linguistics (NAACL), p1937-1946, 2022 (CCF B类)

12. Bo Chen*, Jiayi Liu*†, Mieradilijiang Maimaiti, Xing Gao and Ji Zhang, “Generating Persuasive Responses to User Reviews with Multi-Source Textual Knowledge in E-commerce”, In International Conference on Information and Knowledge Management (CIKM), p2994-3002, 2022 (CCF B类)
13. Yuanhang Zheng, Zhixing Tan, Meng Zhang, Mieradilijiang Maimaiti, Huanbo Luan, Maosong Sun, Qun Liu and Yang Liu, “Self-Supervised Quality Estimation for Machine Translation”, In International Conference on Empirical Methods in Natural Language Processing (EMNLP), p2068-2077, 2021 (CCF B类)
14. Turghun Tayir, Lin Li, Xiaohui Tao, Mieradilijiang Maimaiti, Ming Li, Jianquan Liu, “Visual Pivoting Un-supervised Multimodal Machine Translation in Low-Resource Distant Language Pairs”, In International Conference on Empirical Methods in Natural Language Processing (EMNLP), 2024 (CCF B类)
15. Turghun Tayir, Mieradilijiang Maimaiti, Jianquan Liu, “Low-Resource Machine Translation with Different Granularity Image Features”, In The 7th Chinese Conference on Pattern Recognition and Computer Vision (PRCV), 2024 (CCF C类)
16. Hongbin Na, Zimu Wang, Mieradilijiang Maimaiti*, Tong Chen, Wei Wang, Tao Shen, Ling Chen, “Re-thinking Human-Like Translation Strategy: Integrating Drift-Diffusion Model with Large Language Models”, International Conference of ZZZZZ, 2024 (CCF B类), 通讯作者, 在审
17. 米尔阿迪力江·麦麦提, 吾守尔·斯拉木, 努尔麦麦提·尤鲁瓦斯, 热依曼·吐尔逊, 艾尼宛尔·托乎提, “基于智能手机的维吾尔语语音控制系统的开发” [J]. 计算机应用与软件, 2016.6, 33(06):220-223+305. (北大核心,CSCD)
18. 米尔阿迪力江·麦麦提, 吾守尔·斯拉木, 热依曼·吐尔逊, 迪丽尼格尔·热夏提, “基于Windows Mobile平台的产品管理软件的设计与开发” [J]. 新疆大学学报自然科学版, 2014.11, 31(04):465-469.(CSSCI)
19. 米尔阿迪力江·麦麦提, 吾守尔·斯拉木, 迪丽尼格尔·热夏提, 帕尔哈提·达吾提, “基于智能设备的汉语学习软件的研究与开发” [J].新疆大学学报自然科学版, 2014.5, 31(2):205-211.(CSSCI)
20. 伊·达瓦, 米尔阿迪力江·麦麦提*, “丝绸之路经济带相似语言信息横向处理通信技术的研究” [J].新疆师范大学学报自然科学版, 2014.12,33(4):66-74.(CSSCI), 通讯作者
21. 伊·达瓦, 王羨慧, 米尔阿迪力江·麦麦提, “基于统计机器翻译技术的胡都木一托忒文本转写的实现” [J]. 西部蒙古论坛, 2014.2, 2:62-71.(CSSCI), 通讯作者
22. 迪丽尼格尔·热夏提, 吾守尔·斯拉木, 艾尼宛尔·托乎提, 米尔阿迪力江·麦麦提, “维哈柯文键盘标准制定研究” [J]. 新疆大学学报自然科学版, 2014.2, 31(1):102 -108.(CSSCI)
23. 赵小兵*, 高璐, 高定国, 包乌格德勒, 米尔阿迪力江·麦麦提, 刘洋, 才智杰, 孙媛. “少数民族语言分词技术评测数据集MLWS2021”[J]. 中国科学数据(中英文网络版), 2022, 7(02):6-14. (北大核心,CSCD)
24. 从2014年到2016年发表的文章全列表可以参考我的[个人主页](#).

• 计算机发明专利

– 已授权

- * 米尔阿迪力江·麦麦提, 刘洋, 栾焕博, 孙茂松, “一种基于无监督领域自适应的神经网络机器翻译方法”, 2019.07.12, 专利号: ZL 2017 1 0139214.0;
- * 孙茂松, 米尔阿迪力江·麦麦提, 刘洋, 栾焕博, “神经网络机器翻译模型的训练方法和装置”, 2020.05.19, 专利号: ZL 2018 1 0845896.1;
- * 刘洋, 米尔阿迪力江·麦麦提, 栾焕博, 孙茂松, “基于数据增强的低资源语言神经机器翻译方法与系统”, 2021. 06. 15, 专利号: ZL 2021 1 0857215.5;

– 已受理

- * 米尔阿迪力江·麦麦提, 肖文涛, 周喜, 马玉鹏, 杨雅婷, “基于跨语言表示学习的语义理解方法及装置”, 2024.03.22, 申请号:202410332204.9 (审核中)
- * 米尔阿迪力江·麦麦提, 王晓博, 达吾列·金恩斯别克, “基于多语种机器翻译模型的低资源语言关系抽取数据集构建方法”, 2024.08.21, 申请号:202411148768.3 (审核中)

- 在申请
 - * 李艺飞, 程曦, 米尔阿迪力江·麦麦提, 张太红, “一种基于主动学习的低资源语言机器翻译方法”, 2024.9.17. (代理公司处理中)
- 计算机**软件著作权**
 - 2014-3-11 “维吾尔语多种文字转换软件”, 登记号: 2014SR029071;已发布;
 - 2014-4-8 “基于Android平台的汉语学习软件”, 登记号: 2014SR039439; 已发布;
 - 2014-5-14 “培训机构维吾尔文大规模信息管理软件”, 登记号: 2014SR060362; 已发布;
 - 2014-8-15 “维、哈、柯多语种网络文本采集处理软件”, 登记号: 2014SR147362; 已发布;
 - 2014-9-9 “基于Android平台的维吾尔语语音控制助理软件”, 登记号: 2014SR160254; 已发布;
 - 2016-2-16 “维吾尔语人工词性标注及语料库构建系统”, 登记号: 2016SR031180; 已发布;
 - 2016-3-14 “维吾尔语自动词性标注系统”, 登记号: 2016SR052763; 已发布;
 - 2016-12-19 “维吾尔语自动词干提取与词性标注系统”, 登记号: 2016SR379408; 已发布;
 - 2019-01-30 “基于Python的多语种多文种编码转换工具软件”, 登记号: 2019SR0110291; 已发布;
 - 2019-01-30 “面向低资源语言的多语种在线机器翻译系统”, 登记号: 2019SR0108620; 已发布;
 - 2024-07-22 “基于Transformer的低资源机器翻译软件”, 登记号: 2024SR1037591; 未发布;

学术服务

- 国内学术机构 (专委会) 兼职
 - 中国人工智能学会会员;
 - 中国计算机学会会员;
 - 中国中文信息学会专业会员;
 - 中国人工智能学会多语种智能信息处理专委会执行委员;
 - 中国计算机学会自然语言处理专委会执行委员;
 - 中国中文信息学会**青年工作委员会**会员;
 - 中国中文信息学会民族语言文字信息专委会**常委**及执行委员;
 - 中国中文信息学会机器翻译专委会会员;
 - 中国中文信息学会自然语言生成与智能写作专委会会员;
- 国内学术服务
 - 中国中文信息学会机器翻译大会2024 **赞助主席**;
- 国际会议:
 - **审稿 (Reviewer)**: ACL2016, ACL2023-2024, EACL2024, NAACL2019, NAACL2024, EMNLP2024, AAAI2018, COLING2018, COLING2025, ACM MM2024, IJCNN2022, NLP2024, PACLIC2021, NLP2022, MLIP2022-2023, DMS2023, IEEE CAI2024, Fuzzy2024
 - **技术程序委员会 (TPC)**: AISS2023-2024, AICCC2023, CMCM2024, IEEE MLNLP2024
- 国际期刊 (审稿):
 - International Journal of Intelligent System (IJIS) 2021-2022, ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP) 2022-2023, International Journal of Computer Science Review (JCSR) 2024
- 国际 & 国内重大学术会议报告
 - CIKM 2022, 海报展示, 线上, 2022.10
 - NAACL 2022, 海报展示, 线上, 2022.06
 - EMNLP 2021, **主会口头报告**, 线上, 2021.11
 - ICIS 2018, 主会口头报告, 北京大学, 2018.10
 - ICCSIP 2018, 主会口头报告, 北京大学, 2018.11
 - 第十六届全国少数民族语言文字信息处理学术研讨会, 中国 (桂林), 2017.09

- 国内高校 & 企业学术交流报告
 - “基于未监督领域自适应的低资源语言神经网络机器翻译”, 北京大学, 北京, 2018.09
 - “清华大学博士论坛”, 口头报告, 清华大学, 北京, 2019.3
 - “形态丰富语言之间提高低资源语言神经网络机器翻译模型性能”, 深圳新译科技, 深圳, 2019.6
 - “低资源语言神经网络机器翻译”, 中央民大, 北京, 2019.11.12
 - “阿里巴巴达摩院EMNLP2021见闻”, 杭州, 浙江, 2021.8
 - “浅谈大模型时代下低资源语言机器翻译研究进展”, 伊犁师范大学, 新疆, 2023.11

🔑 专业技能

- 自然语言处理: 机器翻译 (相关技术), 语言模型 (预训练), 信息检索, 文本生成以及文本推理
- 深度学习: 网络架构 (RNN (GRU & LSTM), CNN), 库/工具 (Tensorflow, Pytorch)
- 机器学习: 模型, 策略, 算法, 监督学习, 统计学习, 无监督学习
- 语音信号处理: 语音识别, 语音合成, 口语翻译
- 计算机编程: 基础, 网络开发, 编程设计 (C, C++, C#, Java, Python (机器学习 & 深度学习工具)), 数据库, MVC 架构, 智能移动端
- 其他技能: 平面设计, 项目管理, 组织能力

⚙️ 开发经历

- 团队开发
 - Windows Mobile 产品管理系统, C#, 2009
 - 大型培训学校信息管理平台, C#, 2012
 - 维吾尔语语音合成短信阅读器, Android, 2013
 - 维吾尔语GIS系统, VB, 2013
 - 基于 Moses 与 Microsoft Translator Hub的维吾尔语-汉语统计机器翻译系统, C++/JavaScript/HTML, 2014
 - 汉语语音输入法, Java & Android, 2014
 - 维吾尔语-汉语双向口语翻译系统, C++ & Android, 2014
- 独立开发
 - 图书馆信息管理平台, C#, 2009
 - 高校考试管理系统, C#, 2009
 - 高校寝室信息管理平台, C#, 2010
 - 高校考试监考安排系统, C#, 2011
 - 基于Android的汉语学习平台, Java & Android, 2012
 - 维、哈、柯、蒙、藏多语种在线信息采集与处理平台, C# & ASP.net, 2013
 - 维吾尔语多文中文本转换系统, C#, 2013
 - 维吾尔语词干提取和词性标注预处理工具, C++ & C#, 2014
 - 维吾尔语自动词性标注工具, C++ & C#, 2014
 - 维吾尔语自动词干提取工具, C++ & C#, 2015
 - 维、蒙、藏 - 汉语双向统计机器翻译系统, Python/JavaScript/HTML, 2016
 - 维 - 汉语双向神经机器翻译系统, Python/Tensorflow/C++/JavaScript/HTML, 2016
 - 跨语言信息检索系统, JavaScript/HTML, 2017
 - 低资源语言神经机器翻译系统, Python/Tensorflow /JavaScript/HTML, 2018
 - 多语种神经机器翻译系统, Python/Tensorflow/JavaScript/HTML, 2019

🌟 获奖情况

荣获优秀学生干部	2007 – 2008
荣获优秀共青团员	2008 – 2009
获得国家励志奖学金	2008 – 2009

获得国家励志奖学金	2009 – 2010
荣获大学生志愿者暑期社会实践“先进个人”	2010.10
荣获信息科学与工程学院第二届程序设计大赛“二等奖”	2011.06
获得国家奖学金	2010 – 2011
荣获三好学生标兵	2010 – 2011
荣获新疆大学优秀毕业生	2012.06
荣获新疆大学研究生二等学业奖学金	2012 – 2015
获得国家奖学金	2014 – 2015
获得清华大学研究生社会实践三等奖学金	2016.12
荣获中国电子学会科学技术二等奖	2022.01

♥ 爱好

- 阅读 - 心理学、科幻小说、语言学以及科技类
- 电影 - 武打片 (动作、犯罪)、戏剧、科幻以及美剧, 从不看爱情片
- 体育 - 健身、游泳、台球、爬山、滑雪、踢球、羽毛球 (很少打) 以及户外运动
- 旅游 - 景点、古时建筑、博物馆
- 美食 - 减脂餐、牛排、新疆美食 (手抓羊肉 & 抓饭)、自助餐、不喜欢火锅

🖼️ 国际会议学术报告记录



Figure 1: 国际会议上分享相关科研成果

🖼️ 国内高校学术报告记录

国内高校



清华大学，2019
博士论坛



中央民族大学，2019
论文分享



北京大学，2018
精品电子课程研究中心



新疆大学，2017
机器学习方法分享

一、报告题目一

浅谈大模型时代下低资源机器翻译研究进展

米尔阿迪力江·麦麦提
研究助理

报告时间：2023年11月22日
17:00 - 18:00
腾讯会议ID: 261 877 533

报告人介绍

米尔阿迪力江·麦麦提，博士，中国科学院新疆理化技术研究所助理，曾任阿里巴巴达摩院机器翻译与语言技术实验室 NLP 高级算法工程师，中国人工智能学会多语言智能信息处理专委会委员，中国计算机学会自然语言处理专委会执委以及中国中文信息学会机器翻译专委会委员。曾在多个国际会议（ACL、EMNLP、IJNLP、IJCAI、AAAI、COLING、NAACL、DUNN、ACM TALLP）的审稿人和技术委员会任职。

摘要

近年来，以 ChatGPT 和 GPT-4 为代表的大规模语言模型（简称大模型）展现出了惊人的知识推理和生成能力，吸引了以自然语言处理为核心的 AI 研究和产业社区广泛关注。大模型在自然语言处理领域的应用已经取得了令人瞩目的成就。然而，大模型在资源匮乏的低资源语言翻译方面一直存在挑战，并且在世界多语言翻译任务上带来了重要挑战。本报告从大模型的历史发展、背后的工作原理以及核心技术架构出发，探讨大模型的能力表现，探索自然语言处理研究带来的挑战和机遇以及大模型在机器翻译和大模型微调技术面临的挑战和机遇。此外，报告还将探讨大模型时代下机器翻译的现状，分析多语言大模型在机器翻译中的优势和劣势以及基于大模型的低资源机器翻译研究的前景。最后提出大模型在自然语言处理领域的任务上的几点展望以及该领域在机器翻译研究领域的任务参与开放课题。

中国科学院新疆理化技术研究所
Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences

浅谈大模型时代下低资源机器翻译研究进展

米尔阿迪力江·麦麦提
(伊犁, 2023.11.22)

<http://www.miradelfan.com>

2023/11/23 |

新疆伊犁师范大学，2023
网络安全与信息技术学院

Figure 2: 国内高等院校分享研究成果与调研探索



Figure 3: 给俞正声主席演示自主研发的维吾尔语命令词识别 (语音控制) 系统



Figure 4: 清华大学低资源多语种翻译系统 (小语种)