

# 米尔阿迪力江·麦麦提

✉ miradel\_51@hotmail.com    🏠 <https://www.miradeljan.com>

## 🎓 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

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。

以往的工作经历请参考我的[个人主页](#) 2009 – 2018

## 📄 代表性项目经历

国家自然科学基金项目	2011.03 – 2011.09
“维吾尔文笔记鉴别研究”	

- 地点：新疆大学

- 导师：吾守尔·斯拉木院士
- 贡献：C++开发和系统测试。

#### 国家自然科学基金项目

2012.04 – 2012.09

“面向中西亚的出口软件开发”

- 地点：新疆大学
- 导师：吾守尔·斯拉木院士
- 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

2013.07 – 2013.11

“基于网络的维吾尔文舆情分析大型软件研究”

- 地点：新疆大学
- 导师：吾守尔·斯拉木院士
- 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。

#### 工业与信息化部科技项目

2014.01 – 2014.04

“基于Android的维吾尔语语音控制系统”

- 地点：新疆大学
- 导师：吾守尔·斯拉木院士
- 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。

#### 973项目

2016.11 – 2017.04

“面向三元空间的互联网中文信息处理理论与方法”

- 地点：清华大学
- 导师：孙茂松教授与刘洋教授
- 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

2017.05 – 2018.01

“跨语言社会舆情分析基础理论与关键技术研究”

- 地点：清华大学
- 导师：孙茂松教授与刘洋教授
- 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。

#### 高校-企业合作项目

2018.04 – 2018.09

“基于深度学习的维汉机器翻译”

- 地点：清华大学
- 导师：刘洋教授
- 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。

#### 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”

- 地点：清华大学
- 导师：刘洋教授
- 贡献：主要负责开发低资源神经机器翻译系统。

### • 论文工作

- Bo Chen\*, Jiayi Liu\*<sup>†</sup>, **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, 2022)
- Wenshen Xu\*, **Mieradilijiang Maimaiti**\*, Yuanhang Zheng, Xin Tang, and Ji Zhang<sup>†</sup>, “Auto-MLM: Improved Contrastive Learning for Self-supervised Multi-lingual Knowledge Retrieval”, In Arxiv, 2022
- 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, 2022)
- Xiaobing Zhao, Lu Gao, Dingguo Gao, Wugedele Bao, **Mieradilijiang Maimaiti**, Yang Liu, Zhijie Cai, and Yuan Sun\*, “Minority Languages Word Segmentation Technology Evaluation Dataset MLWS2021”, In China Scientific Data, 2022.
- **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, 2021)
- **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 (ACM TALLIP, 2021)
- **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 System (IJIS, 2021)
- 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, 2021)
- **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 AND TECHNOLOGY (TST, 2020)
- 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, 2020)
- **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 (ACM TALLIP, 2019)
- **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 (ICSSP, 2018)
- **Mieradilijiang Maimaiti**, Xiaohui Zou\*, “Discussion on Bilingual Cognition in International Exchange Activities”, In International Conference on Intelligence Science (ICIS, 2018)
- 从2014年到2016年发表的文章可以参考我的[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”, 计算机发明专利, 授权号: CN 107038159 A; 申请号: 201710139214.0, 已授权;
- \* 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”, 计算机发

明专利，授权号：CN 109117483 B; 申请号：201810845896.1，已授权；

- 审核中

\* 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神经机器翻译方法与系统”，计算机发明专利，申请号：2021108572155，审核中；

#### • 计算机软著

- 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；

## 学术服务

---

#### • 国际会议：

- 审稿：ACL2016, ACL2023, AAI2018, COLING2018, NAACL2019, PACLIC2021, IJCNN2022, NLPAI2022, CVIPPR2023

- 技术程序委员会：ADMIT2022, AMLIT2023, ACDP2023, AIMLR2023, CSSE2023, AISS2023, MLAMDA2023, AACIP2023

#### • 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

## 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)、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、羽毛球以及户外运动
- 旅游 - 景点、古时建筑、博物馆

## 🖥 演示系统

---



Figure 1: 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