

# 张龙辉

✉ longhuizhang97@gmail.com · ☎ 15040150349 · 男 · 28 岁 · 河南

## 🎓 教育背景

哈尔滨工业大学（深圳），广东	2022.09 – 2026.06
在读博士研究生 计算机科学技术，导师：李晶教授（海外优青）、张民教授（国家杰青）	
东北大学，辽宁	2019.09 – 2022.06
硕士研究生 计算机技术，导师：任飞亮教授（小牛翻译实验室）	
沈阳化工大学，辽宁	2015.09 – 2019.06
学士 计算机科学与技术	

## 🔬 研究方向

- 代码大模型，信息检索，信息抽取

## 💼 实习经历

阿里通义实验室，浙江，面向大模型的信息检索方向研究型实习生	2022.09 – 2024.06
<ul style="list-style-type: none"><li>论文: ACL 2024 论文一篇（CCF A 类会议，大模型文本排序方向）、TOIS 论文一篇（CCF A 类期刊，大模型文本-代码混合排序方向）、COLING 2024 一篇（CCF B 类会议，语言模型预训练方向）。</li><li>竞赛: 国际信息检索竞赛 TREC 2022 文本检索任务第一名。</li><li>应用: ACL 2024 产出的文本排序模型 RankingGPT、COLING 2024 产出的中文预训练语言模型 BABERT 已成功部署于阿里云多个重要产品线。</li></ul>	

## ⚙️ 项目经历

Verilog 代码大模型及编程插件研发（学生负责人，华为-哈工大合作项目）	2024.06 – 2025.06
安卓到鸿蒙的工程级应用移植（学生负责人，华为-哈工大合作项目）	2023.06 – 2024.06

## 📖 学术成果

### Progressive Adaptation of Large Language Models for Multilingual Text Ranking

- 期刊: TOIS (CCF A 类期刊, 影响因子 9.1, 一作)
- 研究方向: 代码-自然语言检索
- 论文内容: 针对当前代码-自然语言混合重排序研究匮乏的问题，提出利用大模型进行高质量的多语言重排序数据合成，并通过模型蒸馏策略缓解多语言冲突问题。

### Function-to-Style Guidance of LLMs for Code Translation

- 会议: ICML 2025 (CCF A 类会议, 一作)
- 研究方向: 代码翻译
- 论文内容: 针对当前代码大模型在代码翻译任务上存在正确性低下和风格糟糕的问题，提出了从功能学习到风格学习的两阶段大模型训练策略。

### Speed Up Your Code: Progressive Code Acceleration Through Bidirectional Tree Editing

- 会议: ACL 2025 (CCF A 类会议, 一作)
- 研究方向: 代码优化
- 论文内容: 针对当前代码大模型在代码优化任务上存在的优化路径不清晰的问题，提出通过双向树编辑策略构建高质量指令微调的数据，并通过课程学习提高模型在代码优化任务上的表现。

**A Two-Stage Adaptation of Large Language Models for Text Ranking**

- 会议: ACL 2024 Findings (CCF A 类会议, 一作)
- 研究方向: 自然语言检索
- 论文内容: 针对当前大模型在文本重排序任务上探索不足的问题, 提出了两阶段的训练策略, 通过弱监督数据预训练和有监督数据上的微调, 增强大模型的信息检索能力。
- 成果应用: 研究成果已成功部署于阿里云多个重要产品线, 包括但不限于阿里云的“百炼”与“灵积”大模型平台、智能搜索引擎、NLP 服务平台, 有效赋能了通义 Qwen APP、通义智文、通义听悟、通义星辰等一系列企业创新应用。

**Correct and Efficient Code Translation via Multi-Perspective Exploration and Difference-Aware Selection**

- 会议: ACL 2026 (CCF A 类会议, 一作, 投稿中)
- 研究方向: 代码翻译
- 论文内容: 针对当前的代码翻译系统生成的代码的运行效率普遍低于人类代码的问题, 提出了由“基于并行上下文学习的重复采样”和“diff 感知的大模型判别”组成的测试时扩展框架。

**Chinese Sequence Labeling with Semi-Supervised Boundary-Aware Language Model Pre-training**

- 会议: COLING 2024 (CCF B 类会议, 一作)
- 研究方向: 语言模型的预训练
- 论文内容: 针对语言模型在中文场景下能力受限的问题, 通过半监督学习策略将中文词汇的边界信息注入预训练模型, 强化模型的中文理解能力。

**Hybrid Retrieval and Multi-stage Text Ranking Solution at TREC 2022 Deep Learning Track**

- 会议: TREC 2022
- 研究方向: 信息检索

**A Simple but Effective Bidirectional Framework for Relational Triple Extraction**

- 会议: WSDM 2022 (清华大学计算机推荐 A 类会议, CCF B 类会议, 学生一作)
- 研究方向: 信息抽取

**A Novel Global Feature-Oriented Relational Triple Extraction Model based on Table Filling**

- 会议: EMNLP 2021 (CCF B 类会议, 学生一作)
- 研究方向: 信息抽取
- 学术影响: 谷歌引用 100+, 2022 CCF 大数据与计算智能大赛 (CCF BDCI) “工业知识图谱关系抽取任务”官方基线。

**A Conditional Cascade Model for Relational Triple Extraction**

- 会议: CIKM 2021 (CCF B 类会议, 学生一作)
- 研究方向: 信息抽取

**BSLRel: 基于二元序列标注的级联关系三元组抽取模型**

- 期刊: 中文信息学报 (CCF T1 类期刊, 一作)
- 研究方向: 信息抽取

**🏆 荣誉及比赛经历**

辽宁省优秀硕士学位论文	2022
国际信息检索竞赛 TREC 2022	2022.06 – 2022.09
• 排名: 文本检索任务第一名	
2021 语言与智能技术竞赛: 多形态信息抽取任务, 队长	2021.03 – 2021.05
• 排名: 4/2148, 三等奖	

**2020 语言与智能技术竞赛：关系抽取任务, 队长**

2020.03 – 2020.05

- 排名: 5/1166, 优胜奖