# 王学康

**邮箱：wangxk0223@gmail.com | 电话：+86-13161189787**

## 教育背景

**清华大学 博士一年级（保送直博）** 09/2026-至今

- **专业：计算机科学与技术**
- **导师：王晓智**
- **语言：**英语熟练（CET-4: 617, CET-6: 578，**IELTS: 7.5**）
- **编程能力：ACM-ICPC 银牌**，熟练使用 Python、C++，熟悉 Pytorch 框架下的深度学习算法实现

**北京理工大学 本科** 09/2022-07/2026

- **专业：软件工程**
- **GPA：3.72/4.0 (90.0/100), 排名 6/109**

## 论文

- **Xuekang Wang***, Zhuoyuan Hao*, Shuo Hou, Hao Peng, Juanzi Li, Xiaozhi Wang. Reproducing, Analyzing, and Detecting Reward Hacking in Rubric-Based Reinforcement Learning.（*Under Review, submitted to ACL 2027(ARR August), arXiv:2606.04923*）
- **Xuekang Wang**, Shengyu Zhu, Xueqi Cheng. Speculative Safety-Aware Decoding.（***Accepted, EMNLP 2025 Main Conference***，*arXiv:2508.17739*）
- Yifei Zhu, Hengyu Zhao, Zhongxiang Lei, **Xuekang Wang**, Di Wang, Jinyan Liu. Enhancing Robustness and Privacy: Defense Against Parallel Mixed Threats in Federated Learning.（*Under Review*）
- Ruibiao Fu*, **Xuekang Wang***, April Villeda Roblero, Yang Song. Chatbots on a Mid-scale Knowledge Base: Using an R2 University as an Example.（***Accepted, 2024 ICBAIE***, *Co-first author*）

## 研究经历

**CHERRL:基于准则的强化学习中奖励欺骗现象的可控复现、分析与检测** 12/2024-05/2025

**指导老师：王晓智助理教授，清华大学深圳国际研究生院**

- **发现问题：**阅读 Rubric-based RL 相关论文，发现当前 Rubric-based RL 中经常出现 Reward Hacking 现象，其来自 judge model 内部的多种偏见，难以稳定复现并开展缓解此类现象方法的研究。
- **提出方法：**设计 CHERRL 环境，通过主动注入特定类型的 biased reward 到 proxy reward 中，实现对 policy model Reward Hacking 行为的动态观测与可控复现，并基于该环境研究了不同偏见驱动的 reward hacking 的异同，最后设计了一个自动化检测 reward hacking 的 agent 并在 CHERRL 上验证了其相比于通用 agent 更加优越的准确性和成本消耗。
- **完成实验：**在一台 8 卡 H100 服务器上基于 verl 框架完成了 CHERRL 的实现。在指令遵循、医疗问答和写作任务上复现并研究了 Rubric-based RL 中包括自我褒奖、冗长回复在内的 4 种裁判偏见驱动的 reward hacking 现象。使用训练的 log 数据完成分析部分的实验，并带领其他合作者完成 agent 的设计与评测。
- **任务与成果：**作为第一作者投稿文章到 ARR August（目标会议为 ACL 2027），负责方法提出、全部强化学习部分的实验、agent 部分实验设计、论文撰写与制图制表。

**Speculative Safety-Aware Decoding:推理高效的 LLM 解码阶段安全对齐方法** 12/2024-05/2025

**指导老师：朱胜宇副研究员，中国科学院计算技术研究所**

- **发现问题：**阅读大模型安全对齐和高效推理相关论文，在复现 Deep-Align、SafeDecoding 等方法的开源代码和文献汇报的过程中注意到传统的安全对齐方法往往需要大量的计算资源，如能在推理阶段实现安全防御，可能在节省计算资源和安全对齐的可迁移性上比传统安全对齐方法更具优势。
- **提出方法：**在学习 LLM Safety 相关研究的同时广泛关注其他领域，对 LLM 推理加速的投机解码技术进行了研究，意识到基于小模型的投机解码加速方法可以在调整后应用到解码阶段防御中，提出了利用深度安全对齐的小模型对未深度对齐大模型进行投机解码时的预测接受率实现快速可迁移的安全对齐算法的 idea。
- **完成实验：**编写实验代码，基于 Pytorch 框架和 Transformers 库进行多次 pilot study，在实验中根据观察到的现象逐渐完善算法的细节，最终验证了 idea 的可行性并完成了算法的设计。在 GCG、PAIR 等攻击方法上

**王学康**

**邮箱：wangxk0223@gmail.com | 电话：+86-13161189787**

测试了方法的 safety，在 GSM8K 和 JustEval 数据集上测试了方法的 utility，通过计算 TPS 测试了方法的解码效率，并分别与现有的多种防御的 baseline 进行比较。

- **任务与成果：**作为第一作者投稿 EMNLP 2025，负责方法提出、全部实验、论文实验部分撰写与制图制表。

**FedAR-BPM：一种联邦学习中针对模型混合攻击和隐私攻击的防御机制** 04/2024-12/2024

**指导老师：刘金艳副教授，北京理工大学计算机学院**

- **文献阅读：**阅读相关文献，系统研究了联邦学习中的多种模型攻击（如模型中毒、后门植入等）在混合情景下的防御策略。
- **完成实验：**复现了基于梯度反转攻击的隐私泄露模型，复现并比较了 Laplace、Gauss 等**四种差分隐私防御机制**抑制隐私泄露风险的效果。
- 参与设计和实现 FedAR-BPM 框架，实现了提升联邦学习在拜占庭攻击与隐私攻击场景中的鲁棒性的目的，通过 CIFAR-10 等数据集验证效果。
- **任务与成果：**负责了大部分实验的完成。参与论文的写作过程，汇总实验结果并负责了实验部分的撰写。

**通过基于知识库的大语言模型开发特定领域个性化 ChatBot 的研究** 01/2024-07/2024

**指导老师：Yang Song 教授，北卡罗来那州立大学 2024 冬季 GEARS 科研项目**

- **文献阅读：**使用 Octopus 等自动化爬虫工具抓取数据并构建了约 150,000 tokens 的数据集，为开发 ChatBot 提供数据。
- **完成实验：**设计针对知识库集成的提示词策略，结合 Retrieval-Augmented Generation (RAG) 技术，将特定领域知识注入大语言模型，在多个基础大语言模型上成功构建个性化 ChatBot。
- **任务与成果：**设计了一套系统化的大模型回复质量评估标准，基于该标准开发测试集，并对多个主流大语言模型进行对比测试，提供全面性能分析，最后作为第一作者将研究成果发表于 ICBAIE 2024 会议。

## 荣誉及奖励

- ACM-ICPC 国际大学生程序设计竞赛亚洲区域赛（上海），银牌
- CCPC 中国大学生程序设计竞赛全国邀请赛，银牌
- CCPC 中国大学生程序设计竞赛（济南站），铜牌
- 字节跳动 2024 Byte AI 安全挑战赛，全国决赛第八名（唯一入围决赛本科生队伍）
- 华为 2024 年软件精英挑战赛，京津东北赛区第二名，全国决赛第九名（本科生队伍最好成绩）