陆宇暄 Yuxuan Lu

邮箱: yx_lu@pku.edu.cn
主页: luy.me

个人简介

浙江杭州人，2000 年 7 月生，北京大学计算机学院博士研究生。研究方向为算法博弈论理论及其应用，重点采用信息提取、机制设计与校准理论等手段，系统研究人工智能系统（尤其是语言模型）的评价、激励与行为机制等应用。在严谨理论框架下，针对人工智能系统，探讨其构建路径、脆弱性边界及其优化方案。

教育背景

北京大学 计算机学院 博士研究生 2022.9 - 2027.6（预计）

计算机科学与技术 博士研究生，导师：孔雨晴 长聘副教授

研究方向：算法博弈论，计算经济学

北京大学 信息科学技术学院 本科 2018.9 - 2022.6

计算机科学与技术 学士（图灵班）（Summa Cum Laude）

学术经历

西北大学（Northwestern University）访问博士前研究员 2025.2 - 2025.6

合作导师：Jason Hartline 教授

杜克大学（Duke University）学术实习生 2021.6 - 2021.12

合作导师：Fan Zhang 教授

论文发表 （* 表示共同第一作者）

1. **信息提取 × 大语言模型**

Eliciting Informative Text Evaluations with Large Language Models. *In EC 2024*

Yuxuan Lu*, Shengwei Xu*, Yichi Zhang, Yuqing Kong, Grant Schoenebeck

- 设计基于同伴预测的机制，激励生成高信息量的高自由度文本评价.

Benchmarking LLMs' judgments with no gold standard. *In ICLR 2025*

Shengwei Xu*, Yuxuan Lu*, Grant Schoenebeck, Yuqing Kong

- 提出基于互信息的评测方法，在无真实标签情况下衡量语言模型的回复质量.

Aligned textual scoring rules. *arXiv preprint 2507.06221*

Yuxuan Lu, Yifan Wu, Jason Hartline, Michael J. Curry

- 构建与语言模型交互的适当评分规则，以激励诚实文本化报告.

2. **校准理论 × 人工智能**

Jailbreaking LLMs via Calibration. *arXiv preprint 2602.00619*

Yuxuan Lu, Yongkang Guo, Yuqing Kong

- 将语言模型的越狱问题刻画为再校准过程，并提出基于梯度偏移的攻击方法论以实现越狱.

Making and Evaluating Calibrated Forecasts *arXiv preprint 2510.06388*

Yuxuan Lu, Yifan Wu, Jason Hartline, Lunjia Hu

- 设计一类具有诚实性质的校准指标，以正确的方式评估人工智能模型的概率预测校准程度.

Calibrating "Cheap Signals" in Peer Review without a Prior. *In NeurIPS 2023*

Yuxuan Lu, Yuqing Kong

- 提出无需先验数据的一次性机制，用于校准存在系统偏差的同行评审.

3. 博弈论 × 娱乐

SURPRISE! and When to Schedule It. *In IJCAI 2021*

Zhihuan Huang, Shengwei Xu, You Shan, Yuxuan Lu, Yuqing Kong, Tracy Xiao Liu, Grant Schoenebeck

- 经验性分析电竞比赛中"惊喜"出现时机对观众体验质量的影响.

How Gold to Make the Golden Snitch: Designing the "Game Changer" in Esports. *arXiv preprint 2405.19843*

Zhihuan Huang*, Yuxuan Lu*, Yongkang Guo, Yuqing Kong

- 同时从理论与实证角度分析电子游戏中关键转折机制的奖励设计问题.

4. 区块链

FileInsurer: A Scalable and Reliable Protocol for Decentralized File Storage in Blockchain. *In ICDCS 2022*

Hongyin Chen*, Yuxuan Lu*, Yukun Cheng

- 提出去中心化文件存储协议，以提升区块链系统的可扩展性与可靠性.

Empirical Analysis of EIP-1559: Transaction Fees, Waiting Time, and Consensus Security. *In CCS 2022*

Yulin Liu, Yuxuan Lu, Kartik Nayak, Fan Zhang, Luyao Zhang, Yinhong Zhao

- 对 EIP-1559 交易费机制的实际影响进行系统性实证分析.

A Framework of Transaction Packaging in High-throughput Blockchains. *arXiv preprint 2301.10944*

Yuxuan Lu, Qian Qi, Xi Chen

- 构建高吞吐区块链中交易打包博弈的博弈论分析框架以分析矿工潜在的策略性行为.

学术服务

程序委员会委员 (Program Committee Member), ACM EC 2026

Workshop 组织者, WINE 2024

审稿人 (Reviewer), NeurIPS 2025; ICLR 2026; ICML 2026

技能

AI 编程工具: 熟悉 AI 编程工具的优势与性能边界，目前已在相关项目中深度使用 AI 编程工具

编程语言: 熟练使用 Python, C, C++; 熟悉 MATLAB, Mathematica, Solidity, HTML

英语: TOEFL R: 30 L: 30 S: 24 W: 28 Total: 112

荣誉与奖励

北京大学比亚迪奖学金	2024.9
北京大学 John Hopcroft 奖学金	2021.9
2020 年 ACM-ICPC 亚洲区预选赛上海站金牌	2020.12
北京大学二等奖学金	2020.9
2019 年 ACM-ICPC 亚洲区预选赛上海站金牌	2020.12
2019 年 ACM-ICPC 亚洲区预选赛沈阳站金牌	2020.12
北京大学三等奖学金	2020.9
2018 年 ACM-ICPC 亚洲区预选赛沈阳站金牌	2020.12
北京大学新生奖学金	2018.9
第 34 届全国青少年信息学奥林匹克竞赛 (NOI) 金牌	2017.7
第 33 届全国青少年信息学奥林匹克竞赛 (NOI) 金牌	2016.7