Yuxuan Lu

E-mail: yx_lu@pku.edu.cn
Webpage: luy.me

Personal Statement

I am a fourth-year Ph.D. student at Peking University, specializing in the application of game-theoretic principles to artificial intelligence. My research explores how tools from information elicitation, mechanism design, and calibration theory can be used to build, destroy, or improve AI systems, particularly in the context of large language models.

I am driven by the goal of bridging rigorous theoretical foundations with practical AI deployment, developing mechanisms that are not only provably sound but also effective in real-world settings.

Education

Peking University, School of Computer Science Sep 2022 – Jun 2027 (expected)

Ph.D. candidate, advised by Prof. Yuqing Kong

Research Directions: Game Theory, Information Elicitation, Large Language Models

Peking University, School of EECS Sep 2018 – Jun 2022

B.S. in Computer Science, Summa Cum Laude, Member of Turing Class

Experience

Visiting Predoctoral Fellow, Northwestern University Feb 2025 – Jun 2025

Hosted by Prof. Jason Hartline

Teaching Assistant, Peking University

Course: Mathematical Foundations for the Information Age (2021 Fall, 2022 Fall)

Course: Algorithm Design and Analysis (2021 Spring)

Research Intern, Duke University Jun 2021 – Dec 2021

Hosted by Prof. Fan Zhang

Selected Publications (* indicates equal contribution.)

Main research topic: Game Theory for AI

1. Information Elicitation + Large Language Models

Eliciting Informative Text Evaluations with Large Language Models. *In EC'24*

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

- Design a peer prediction mechanism to incentivize informative free-form text evaluations from LLMs.

Benchmarking LLMs' judgments with no gold standard. *In ICLR'25*

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

- Propose a mutual-information-based metric to benchmark LLM response without ground truth.

Aligned textual scoring rules. *arXiv preprint 2507.06221*

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

- Develop LLM embedded proper scoring rules to incentivize truthful textual judgments.

2. Calibration + AI

Jailbreaking LLMs via Calibration. *arXiv preprint 2602.00619*

Yuxuan Lu, Yongkang Guo, Yuqing Kong

- Reframe LLM jailbreaking as recalibration and propose a gradient-shift attack to bypass safety guardrails.

Making and Evaluating Calibrated Forecasts *arXiv preprint 2510.06388*

Yuxuan Lu, Yifan Wu, Jason Hartline, Lunjia Hu

- Design a class of truthful calibration metrics for evaluating AI model forecasts in a "correct" way.

Calibrating "Cheap Signals" in Peer Review without a Prior. *In NeurIPS'23*
Yuxuan Lu, Yuqing Kong
- Propose a one-shot mechanism to calibrate biased peer review without prior data.

Side topics: Blockchain, Computation Economics, etc.

1. Blockchain

FileInsurer: A Scalable and Reliable Protocol for Decentralized File Storage in Blockchain. *In ICDCS'22*
Hongyin Chen*, Yuxuan Lu*, Yukun Cheng
- Propose a decentralized file storage protocol in blockchain to enhance both scalability and reliability.

Empirical Analysis of EIP-1559: Transaction Fees, Waiting Time, and Consensus Security. *In CCS'22*
Yulin Liu, Yuxuan Lu, Kartik Nayak, Fan Zhang, Luyao Zhang, Yinhong Zhao
- Conduct an empirical evaluation of the real-world impact of Ethereum's EIP-1559 transaction fee mechanism.

A Framework of Transaction Packaging in High-throughput Blockchains. *arXiv preprint 2301.10944*
Yuxuan Lu, Qian Qi, Xi Chen
- Develop a game-theoretic framework for transaction packaging game in high-throughput blockchains.

2. Game Theory for entertainment

SURPRISE! and When to Schedule It. *In IJCAI'21*
Zhihuan Huang, Shengwei Xu, You Shan, Yuxuan Lu, Yuqing Kong, Tracy Xiao Liu, Grant Schoenebeck
- Quantifies how when surprise occurs during a live esports game influences audience-perceived quality.

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
- Theorize and empirically analyze how to set the reward for a "game changer" item in esports games.

Academic Service

Program Committee Member, EC 2026

Workshop Organizer, WINE 2024

Reviewer, NeurIPS 2025; ICLR 2026; ICML 2026

Skills

AI tools Skills: Familiar with AI coding tools and use them extensively in projects

Programming Skills: Proficiency in Python, C, C++; Familiar with MATLAB, Mathematica, Solidity, and HTML

Language Skills: English: TOEFL R: 30 L: 30 S: 24 W: 28 Total: 112

Awards and Honors

BYD Scholarship of Peking University	Sep 2024
John Hopcroft Scholarship of Peking University	Sep 2021
Gold Medal in 2020 ACM-ICPC Asia Regional Contest Shanghai Site	Dec 2020
Second-class Scholarship of Peking University	Sep 2020
Gold Medal in 2019 ACM-ICPC Asia Regional Contest Shanghai Site	Dec 2019
Gold Medal in 2019 ACM-ICPC Asia Regional Contest Shenyang Site	Dec 2019
Third-class Scholarship of Peking University	Sep 2019
Gold Medal in 2018 ACM-ICPC Asia Regional Contest Shenyang Site	Dec 2018
Freshman scholarship of Peking University	Sep 2018
Gold Medal in the 34th National Olympiad in Informatics	Jul 2017
Gold Medal in the 33th National Olympiad in Informatics	Jul 2016