Bufan(Cici) Gao

github.com/jouisseuse · bufan@uchicago.edu · lkcici4@gmail.com

1. Research Interests

Multi-agent Systems in Artificial Intelligence and Humans; Bias in Computational Systems; AI Safety and Fairness; Computational Social Science; Human-AI Collaboration

2. Education

The University of Chicago 2025 — Present
PhD Student in Social Psychology, advised by Prof. Xuechunzi Bai

Fudan University 2022 — 2025
Master in Computer Science, advised by Prof. Yuan Zhang and Prof. Min Yang

Fudan University 2018 — 2022
B.S. in Department of Computer Science, majored in Information Security

3. Publications

Published:

- **SCTrans: Constructing a Large Public Scenario Dataset for Simulation Testing of Autonomous Driving Systems**
  Jiarun Dai*, **Bufan Gao***, Mingyuan Luo, Zongan Huang, Zhongrui Li, Yuan Zhang, Min Yang
  In Proceedings of the 2024 International Conference on Software Engineering (ICSE'24)

In Preparation:

- **Microfoundations of Statistical Discrimination**
  Bufan Gao, Xuechunzi Bai
  Accepted for a parallel talk in IC2S2 2025
- **Measuring Bias or Measuring the Task: Understanding the Brittle Nature of LLM Gender Biases**
  Bufan Gao, Elisa Kreiss
  Submitted to EMNLP 2025
- **ExplorIt: Facilitating Simulation-based Fuzzing for Autonomous Driving Systems through Multi-Agent Interaction Modeling**
  Bufan Gao, Jiarun Dai, Zongan Huang, Kairui Yang, Yixing Luo, Yingjie Fu, Yuan Zhang, Min Yang, Tao Xie
  Submitted to the Proceedings of the 2026 International Conference on Software Engineering (ICSE'26)

4. Research Experiences

2024 - Present **Research Assistant, University of Chicago**
Bai Lab
Working with Prof. Xuechunzi Bai, I explored the origin of statistical discrimination, a key driver of group disparities in decision-making. Prior models, like the self-fulfilling prophecy, attribute disparities to feedback loops that influence individual productivity over time, breaking group symmetry. My work introduces a new explanation: information sharing. Using the explore-exploit framework in reinforcement learning, we showed that information sharing initially reduces disparities through increasing exploration but later amplifies them via consensus formation and information homogenization. Large language models (LLMs) may intensify consensus effects and accelerate the formation of disparities.

To further explore this mechanism, we plan to conduct human experiments to validate its impact in real-world decision-making contexts.

2024 - Present **Research Assistant, UCLA**
Coalas Lab
While working with Prof. Elisa Kreiss, I develop reliable measurements for bias in large language models (LLMs). Our research highlights the fragmented nature of existing bias assessments, which are often tailored to specific tasks, metrics, or scenarios, resulting in inconsistent and limited perspectives. To address this, we design a comprehensive measurement framework that integrates key evaluation variables and systematically tests bias across diverse conditions. This framework focuses on improving consistency and reliability, providing a robust foundation for bias assessment in LLMs.

2021 - Present **B.S. and Master Student, Fudan University**
Secsys Lab
Under the supervision of Prof. Min Yang and Prof. Yuan Zhang, with the focus on the ***Safety and Robustness of Autonomous Driving Systems (ADS)***, I am working on (1) *Automated Scenario Dataset Construction*: We propose a transformation-based approach, SCTrans, to construct simulation scenario files compatible with the most advanced ADS simulation platforms. (2) *Intelligent NPC Behavior Controller*: Led the design of an NPC behavior controller for the ADS simulation tests, utilizing game theory and social-oriented values to identify ADS-primary responsible accidents. (3) *LLM-driven Scenario Generation*: Generating diverse safety-critical scenarios in high-dimensional spaces using descriptions to drive LLM-based scenario creation (ongoing). (4) *Multi-agent Interaction Modeling ADS Fuzzer*: To enhance ADS safety assessment, we developed ExplorIt, a fuzzer that uses multi-agent interaction modeling to guide scenario mutations and scheduling, improving the detection of unique, ADS-responsible accidents (submitted).

2021 Summer **Computer Science Summer Research Fellowship Program, University of Chicago**
Advised by Prof. Junchen Jiang and Prof. Nick Feamster, I worked on *ADS Fuzzing*: For the safety assessment of autonomous driving systems (ADS), scenario-based simulation testing has become an important complementary technique to physical road testing. To improve efficiency, i.e. trigger more ADS accidents, we integrated fuzzing technique into scenario-based ADS simulation test to mutate more safety-critical scenarios.

5. Honors & Achievements

Venustech Scholarship (Top 3% cross department), 2023
Champion of 2022 DEFCON AutoDriving Contest, 2022
Fudan University Freshman Scholarship, 2022
Champion of 2021 DEFCON AutoDriving Contest, 2021
Fudan University Liu Yong Ling Foundation Scholarship (Top 5% cross university), 2019

6. Service

ICSE 2025 workshop Committee Member

7. Teaching

Principles of Reverse Engineering (PoRE), Teaching Assistant Spring 2023, Spring 2022

8. Skills

Python; Java; C/C++; Prompt Engineering; LLVM; Reverse Engineering; Docker