

# ZIHE (ZOE) SONG

UT Dallas ♦ +1 (469) 900-6071 ♦ [zihe.song@utdallas.edu](mailto:zihe.song@utdallas.edu)

## EDUCATION

### University of Texas at Dallas

2020 - Present

#### Ph.D. Candidate in Computer Science

- Advisor: Dr. Wei Yang
- Specializing in Android systems, program analysis, ML infrastructure, and large-scale debugging and testing tools.
- Experienced in building production-grade systems across analysis, UI exploration, and ML evaluation pipelines.
- Published at ISSTA, USENIX Security, CVPR, ASPLOS, ICSE, MobileSoft, etc. Finalist in Amazon Nova AI Challenge.

### University of Texas at Dallas

2018 - 2020

#### M.S. in Computer Science

### University of Electronic Science and Technology of China

2014 - 2018

#### B.E. in Communication Engineering

## SELECTED PUBLICATIONS

(FULL PUBLICATION LIST AVAILABLE ON [GOOGLE SCHOLAR](#))

- ❖ An Investigation on Numerical Bugs in GPU Programs Towards Automated Bug Detection. *ISSTA 2025*.
- ❖ SoK: Efficiency Robustness of Dynamic Deep Learning Systems. *USENIX Security Symposium 2025*.
- ❖ Can you mimic me? Exploring the Use of Android Record & Replay Tools in Debugging. *MobileSoft 2025*.
- ❖ TAOPT: Tool-Agnostic Optimization of Parallelized Automated Mobile UI Testing. *ASPLOS 2025*.
- ❖ Guardian: A Runtime Framework for LLM-Based UI Exploration. *ISSTA 2024*.
- ❖ WEFix: Intelligent Automatic Generation of Explicit Waits for Efficient Web End-to-End Flaky Tests. *WWW 2024*.
- ❖ An Empirical Analysis of Compatibility Issues for Industrial Mobile Games. *ISSRE 2022*.
- ❖ NICGSlowDown: Evaluating the Efficiency Robustness of Neural Image Caption Generation Models. *CVPR 2022*.
- ❖ NMTSloth: Understanding and Testing Efficiency Degradation of Neural Machine Translation Systems. *ESEC/FSE 2022*.

## ONGOING PROJECT

### Multimodal Agent Framework for Android App Generation (ICLR 2026 under review)

- Built a benchmark of 101 real-world Android tasks requiring LLMs to implement full apps from natural-language specs.
- Developed a multi-agent system to summarize app docs and navigate apps to generate functional test cases automatically.
- Designed an automated evaluation pipeline revealing that current LLMs achieve only ~18.8% functional correctness.

### LLM-Driven Smartphone Interaction Enhancement

- Built an LLM-driven interaction framework that improves intent disambiguation and UI navigation reliability across apps.
- Built a multi-agent reasoning pipeline combining screenshots, view hierarchies, and NL instructions for robust control.
- Designed clarification-prompt models that reduce ambiguous user intents and increase UI navigation task success rates.

## INTERNSHIP / COMPETITIONS

### Amazon Nova AI Challenge - Finalist (Published in Amazon Science Trusted AI 2025)

- Designed adversarial attack strategies for LLMs, uncovering multiple high-severity vulnerabilities in generation behavior.
- Contributed to COMET, a malicious-elicitation framework that improved evaluation throughput and automated red teaming.

### Research Intern - NetEase Fuxi Lab, 05/2020 - 12/2020

- Built an automated game testing framework to simulate diverse player behaviors and increase overall interaction coverage.
- Applied GAIL-based RL models to generate gameplay trajectories, reducing QA workload and improving test robustness.
- Conducted a large-scale study on game compatibility issues, informing debugging workflows and device optimizations.

### Graduate Research Supervision

- Advised master's students on software engineering research projects, an ongoing submission now under review at TOSEM.

## TECHNICAL SKILLS

- ❖ Programming: Python, Java, Kotlin, C/C++, Bash
- ❖ Machine Learning / AI: LLMs, Generative AI, NLP, Reinforcement Learning, ML Evaluation, Adversarial Testing
- ❖ Systems & Mobile: Android Framework, ADB, UIAutomator, View Hierarchy Analysis, Performance Profiling, Tracing
- ❖ Program Analysis: Static Analysis, Dynamic Analysis, Debugging Tools, Runtime Monitoring
- ❖ Tools & Frameworks: CI/CD, Test Automation Frameworks, Profiling Pipelines, PyTorch, TensorFlow, Docker, Git, Linux