

# Yuheng Zhu

📍 Raleigh, NC    ✉ yzhu63@ncsu.edu    ☎ (919)818-9960    🌐 yuhengzhu.me    in yzhu27    📺 yzhu27

## Education

<b>North Carolina State University</b> <i>Ph.D. in Computer Science</i>	<i>Aug 2024 – Present</i>
<b>North Carolina State University</b> <i>Master in Computer Science</i>	<i>Aug 2022 – May 2024</i>
<b>Southern University of Science and Technology</b> <i>B.E. in Computer Science and Technology</i>	<i>Sept 2017 – May 2021</i>

## Experience

<b>Engineer Intern</b> <i>Qualcomm</i>	<i>San Diego, CA</i> <i>Jun 2025 - Aug 2025</i>
<b>Research Assistant</b> <i>North Carolina State University</i>	<i>Raleigh, NC</i> <i>Sept 2023 - Present</i>
<b>Research Assistant</b> <i>Southern University of Science and Technology</i>	<i>Shenzhen, China</i> <i>June 2021 – July 2022</i>
<b>Software Engineer Intern</b> <i>JD.com, AI Research Center</i>	<i>Chengdu, China</i> <i>June 2020 – Aug 2020</i>

## Publications

<b>AdaptAV: Continuous Adaption of Vision Models for Autonomous Vehicles Using Cloud-based Oracle.</b> <i>Yuheng Z.</i> , Dhruva U., Boluo G. & Man-Ki Y., <i>Proceedings of the 100th IEEE Vehicular Technology Conference</i>
<b>Bridging Data and Knowledge: A Neurosymbolic Framework for Reliable Network Analysis.</b> Zhjin Y., <i>Yuheng Z.</i> , Mingzhe C. & Yuchen L., <i>Proceedings of 2025 IEEE Global Communications Conference</i> (Accepted)
<b>FrameScope: Temporal Data Valuation for Stream Active Learning in Autonomous Vehicle Systems.</b> <i>Yuheng Z.</i> , & Man-Ki Y. <i>Proceedings of The 10th ACM/IEEE Symposium on Edge Computing</i> (Accepted)

## Projects

<b>Agentic LLM Testcase Triage Framework</b>	<i>Qualcomm, 2025</i>
<ul style="list-style-type: none"><li>Independently built an agentic RAG testing failure analysis pipeline for 10K+ tests/day.</li><li>Multi-source logs distilled via LLM summarization + error-anchor context, then chunked and indexed in a vector storage (ChromaDB).</li><li>Orchestrated with LangChain agents + function calling for hybrid retrieval (semantic + keyword) and top-k reranking, emitting JSON-schema reports.</li><li>Tools used: <b>LangChain, ChromaDB, FAISS</b></li></ul>	
<b>On-the-fly Coding of Vision Inputs for Evidence-Preserving Perception</b>	<i>NCSU, 2023 - 2024</i>
<ul style="list-style-type: none"><li>Designed and implemented a visual perception middleware for autonomous driving systems that preserves complete and reproducible evidence of visual inputs.</li><li>Utilized compression techniques such as JPEG and H.264, with performance optimizations achieved through NVENC and NVDEC hardware acceleration engines.</li><li>Tools used: <b>C, Python, V4L2, , NVIDIA Jetpack, NVIDIA Codec, FFmpeg</b></li></ul>	
<b>Carla1s Distributed Online Simulation Platform</b>	<i>SUSTech, 2021 - 2022</i>
<ul style="list-style-type: none"><li>Designed and developed a Carla-based remote and hardware-free web simulation platform.</li><li>Developed a <b>backend streaming server</b> for <b>HTTP Streaming ROS</b> Image Topics and Carla Sensor Output, supporting multiple video encodings and streaming protocols, reduce latency of remote simulation from 1-2s to 50ms.</li><li>Tools used: <b>Python Flask, FFmpeg, Ansible, Git, Docker</b></li></ul>	
<b>Jingxiaozi AI Customer Service System</b>	<i>JD.com, 2020</i>
<ul style="list-style-type: none"><li>Participated in two development cycles (2.1.6 &amp; 2.1.7) of JD Jingxiaozi online AI customer service system, made in-depth research on the application and practice of Chinese <b>NLP</b> in e-commerce scenarios</li><li>Developed web API automated testing modules in <b>Java</b> based on JD JSF <b>RPC framework</b>, made test framework can fully mimic user operation logic, increased API test coverage to 100%</li></ul>	

- Tools used: **Java, RESTful API, JavaScript, Redis, RabbitMQ**

## Technologies

---

**Languages:** C, C<sup>++</sup>, Shell, Python, Java, JavaScript, Lua

**Skills:** Linux Kernel, ROS2, Networking, DevOps, Docker, Redis, Carla, ADAS, LangChain, ChromaDB