

CICI (YUTONG) CHENG PH.D.

(+1) 540-666-1970
yutongcheng@vt.edu
ccsnw127.github.io

EDUCATION

Department of Computer Science, Virginia Tech Blacksburg, VA, United States
Ph.D. in Computer Science 2024 - 2028 (*expected*)

- Advisor: Prof. Peng Gao
- Research area: Large Language Model and its Applications, Security, Software Engineering
- Selected courses: Natural Language Processing, System and Software Security, Trustworthiness of Large Language Models

School of Science and Engineering, Chinese Uni. of Hong Kong Shenzhen, China
MPhil in Computer and Information Engineering 2022 - 2024 (*incomplete*)

- Advisor: Prof. Pinjia He
- Research area: Large Language Model and its Applications, Software Engineering
- Selected courses: Deep Learning Foundations and their Applications, Image Processing and Computer Vision, Mobile Computing and Internet of Things

School of Software Engineering, Beijing Jiaotong University Beijing, China
B.E. in Software Engineering 2018 - 2022

- Selected courses: Software System Analysis and Design, Software Architecture, Software Testing Technology and Practice, Computer Network, Database Systems, Reinforcement Learning

PUBLICATIONS

1. **Yutong Cheng**, Osama Bajaber, Peng Gao, Dawn Song. CTINEXUS: Unleash Cyber Threat Intelligence via In-Context Learning of Large Language Models. *Under peer review*, 2024.
2. Xu, Junjielong and Fu, Qiulai and Zhu, Zhouxiang and **Cheng, Yutong** and Li, Zhijiang and Ma, Yuchi and He, Pinjia. Hue: A user-adaptive parser for hybrid logs. *Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE)*, 2023.

INTERNSHIPS & EXPERIENCE

AtomEcho (Startup) | Beijing, China 2023.06 - 2023.12

- Developed a framework (**AtomTool**) for integrating tool utilization skills into large language models (LLMs), enhancing zero-shot generalization for Chinese models.
- Implemented API integrations for advanced tool usage, improving LLM accuracy in handling complex API parameter settings.
- Addressed the hallucination phenomenon by refining prompt design and retrieval enhancements, improving LLM response accuracy and reliability.

Llama Family (Community) | Beijing, China 2023.06 - current

- Fine-tuned Llama2-Chat model using Chinese instruction datasets to enhance conversational abilities in Chinese.
- Integrated LangChain framework into the fine-tuned model for improved performance and functionality in Chinese applications.
- Maintained documentation and updates on GitHub Repo.

PROJECTS	Characterizing, Detecting, and Correcting Comment Errors in Smart Contract Functions via In-Context Learning of Large Language Model <i>Github Repo</i> 2023.06 - 2023.10 This project introduces <i>CETerminator</i> , an automated approach for detecting and correcting Natspec comment errors using the in-context learning capability of large language models.
	CTINEXUS: Unleash Cyber Threat Intelligence via In-Context Learning of Large Language Models. <i>Github Repo</i> 2023.10 - 2024.7 This project introduces <i>CTINexus</i> , a novel framework utilizing in-context learning of large language models for high-quality cybersecurity knowledge graph construction, encompassing structured information extraction, entity alignment, and knowledge reasoning.
AWARDS & HONORS	<ul style="list-style-type: none"> • Bitshares Fellowship, for smart-contract-related research project 2024.09 • Virginia Tech CS Departmental Travel Grant 2024.06 • CCI SWVA Cyber Innovation Scholarship, for CTI-related research project 2024.03
	<ul style="list-style-type: none"> • Lab Lead, for VT Security & Intelligence Lab 2024.08 • Student Organizer, for the 2024 DMV Security Workshop.
	<ul style="list-style-type: none"> • External Reviewer, for NDSS 2025 (Top-tier Conference in Security) 2024.08 • External Reviewer, for ACM CCS 2024 (Top-tier Conference in Security) 2024.03 • External Reviewer, for DLSP 2024 2024.02 • External Reviewer, for TOPS 2023 2024.01 • External Reviewer, for IEEE EuroS&P 2024 2023.12
PROFESSIONAL SERVICES	
SKILLS	Programming: Python, C++/C, Java.
	LLM Framework: LangChain, LlamaIndex, AutoGen, QLoRA.
	Deep Learning Framework: PyTorch, TensorFlow.