Chengjie Lu

+47-413-16-813 / +86-186-0254-3566 | Oslo, Norway | chengjielu@simula.no

https://chengjie-lu.github.io/

Google Scholar | Chengjie-Lu | GitHub

Profile

My research focuses on ensuring the **safety and trustworthiness** of **industrial-scale AI-empowered software systems**, particularly **embodied AI** (e.g., autonomous driving, robotics), and their applications in various domains, including **healthcare** and **industrial disassembly robotics**, through AI-driven validation methods, **digital twin** technologies, **uncertainty** quantification methods, and the use of **foundation models** (e.g., GPT and VLA) to build trustworthy systems.

Education

University of Oslo 06.2022 - Now
Ph.D. in Computer Science, Expected Graduation: 06.2026 Oslo, Norway

Research Areas: Safety & Trustworthiness Assurance of AI-Empowered Software Systems

Nanjing University of Aeronautics and Astronautics 09.2019 - 04.2022
M.Sc. in Computer Science Nanjing, China

Research Areas: AI-empowered Safety Assurance of Autonomous Driving Systems

Nanjing University of Aeronautics and Astronautics 09.2015 - 06.2019
B.Sc. in Software Engineering Nanjing, China

Professional Experience

Simula Research Laboratory, Department of ComplexSE 06.2022 - Present
Ph.D. Research Fellow. I am mainly **involved in three projects**: Oslo, Norway

RoboSapiens (EU Horizon Project) — AI for Embodied Robotics

*This project studies the safety and trustworthiness of self-adaptive robots in unforeseen environments. My contributions include: 1) **supporting trustworthy self-adaptation for industrial-scale robotics in the presence of uncertainties**; 2) developing a **robust self-adaptive framework to support runtime safety and functionality validation**; 3) **applying the research findings across diverse industrial robotic contexts;** see concrete works below*:

- **Assessing the Uncertainty and Robustness of the Industrial Disassembly Robots [4], *ICST2025*** [GitHub]
  - Collaborated with **Danish Technological Institute**, a leading institute in industrial robotics in Denmark. [Web]
  - Supported trustworthy decision-making for industrial disassembly robots by uncertainty and robustness assessment.
  - Provided industrial guidelines that enable robotics to estimate their confidence and adapt their behavior.
- **Vision Language Model (VLA)-Driven Human-Centric Validation for Industrial Robots [6], *ICSE2026*** [GitHub]
  - Collaborated with **PAL Robotics**, a leading robotic company in Spain. [Web]
  - Focused on safety and trustworthiness assurance for human-robot interaction.
  - Leveraged VLM commonsense reasoning to generate realistic human behaviors for robotic validation.
  - Evaluated on industrial-scale warehouse robots, achieving an average of 3.5× improvement in collision detection and 1.63× higher scenario diversity over baselines, demonstrating scalable validation for human-robot interaction.
- **Uncertainty and Quality Evaluation of Vision-Language-Action Model (VLA)-Empowered Robots [14]** [GitHub]
  - Addressed the non-standardized evaluation of VLA models by proposing a systematic framework of quality and uncertainty metrics beyond coarse-grained success rates.
  - Validated across 908 executions on three state-of-the-art VLA models and four robotic tasks, with metrics significantly correlating with expert assessments. Manuscript submitted to IEEE TSE.

AIT4CR (Research Council of Norway Project) — AI for Healthcare

*This is an **interdisciplinary project that develops AI-powered infrastructures to ensure the quality and reliability of Norway's Cancer Registry System**. My contributions include 1) **building AI-powered infrastructures** to support the continuous evolution and reliable scaling of the cancer registry system; 2) **architecting Digital Twins with active transfer learning** to enable safe experimentation without interfering with the live production environment; see concrete work below:*

- **Active Knowledge Transfer for Evolving Cyber-Cyber Digital Twins in Cancer Registry [3], *ESEC/FSE2023*** [GitHub]
  - Collaborated with **Cancer Registry of Norway**, a national institute tracing 1.7 million cancer patients. [Web]
  - Built Cyber-Cyber Digital Twin (CCDT) for the Norwegian cancer registry and employed active knowledge transfer to keep CCDT synchronized with the production system.

◦ Supported safe experimentation and analysis activities without intensively interfering with the live environment.
◦ Ensured the reliability of high-stakes, data-critical infrastructures as they evolve, providing a foundation for trustworthy AI deployment in public societal sectors.

Co-Evolver (Research Council of Norway Project) — AI for Autonomous Driving

This project supports coevolution designs for self-adaptive systems using advanced techniques like evolutionary computation and machine learning. My contributions are: 1) ***supporting safe and trustworthy autonomous driving*** *by developing AI-driven validation methods that efficiently identify critical failure scenarios before deployment; 2)* ***assessing the validity of simulation-based evaluation by leveraging LLMs*** *to assess the realism and quality of generated scenarios; see concrete works below:*

• **Efficient Autonomous Vehicle Validation via Epigenetic and Attention Mechanisms [2], *IEEE TSE, 2022*** [GitHub]
◦ Developed a bio-inspired framework using epigenetic gene-silencing to suppress non-critical factors and concentrate search on high-risk environmental features.
◦ Designed a fusion transformer to extract features from multi-modal environmental data.
◦ Supported safe experimentation and analysis activities without intensively interfering with the live environment.
◦ Validated on **CARLA** with state-of-the-art AV controllers **TCP and Interfuser**, achieving up to 3× improvement in collision detection with faster speed.

• **Adaptive Autonomous Vehicle Validation via Reinforcement Learning [1], *IEEE TSE, 2022*** [GitHub]
◦ Formulated edge-case discovery as a sequential decision-making problem and trained an adversarial RL agent to uncover critical failure scenarios.
◦ Designed a lightweight framework providing 52 environmental configuration REST API endpoints.
◦ Validated on **Baidu Apollo**, where the approach successfully exposed system-level vulnerabilities, achieving 2.6× improvement in collision detection while reducing execution time to 0.73× of baselines.
◦ Released *DeepScenario* benchmark dataset [5] with over 30K driving scenarios to support autonomous driving research. **The dataset was published at MSR 2023**.

• **Assessing Realism of Driving Scenarios with Large Language Models (LLMs) [7], *FORGE2024*** [GitHub]
◦ Evaluated the commonsense reasoning capabilities of LLMs (GPT-3.5, Llama2, Mistral) for assessing the realism of autonomous driving scenarios for real-world deployment.
◦ Proposed novel LLM robustness metrics and benchmarked various LLMs across 576 scenarios with systematic variations in road types, weather conditions, and vehicle parameters.
◦ GPT-3.5 achieved the highest robustness score (12.59/20), demonstrating that LLMs can serve as scalable and cost-effective realism evaluators for sim-to-real transfer.
◦ Findings lay the foundation for bridging the sim-to-real gap in autonomous driving validation using LLM.

Honors and Awards

The Chinese Government Award for Outstanding Self-financed Students Abroad, Group A 2025

Bronze Medal, ACM Student Research Competition Grand Final 2024

Gold Medal, ACM Student Research Competition@ICSE 2023

NUAA Outstanding Graduate & Master Thesis Award 2022

Academic Services

Academic Journal Reviewer

IEEE Transactions on Software Engineering
ACM Transactions on Software Engineering and Methodology
IEEE Transactions on Intelligent Transportation Systems
IEEE Transactions on Industrial Informatics
Science of Computer Programming

Conference Program Committee Member

ISSRE2024 Artifact Evaluation Track
ICST2024 Workshop, ICST2025 Workshop
SE4ADS2025/2026

Supervisory Experience

Co-supervised one master student for his thesis, guiding research design, methodology, analysis, and thesis writing.

Skills and Interests

Languages: English (Fluent); Chinese (Native); Norwegian (A2 Level)
Technical Skills: Python, PyTorch, Tensorflow, LaTeX, ChatGPT/OpenAI APIs
Interests: Oslo Obstacle Race (Completed), Nanjing Marathon (Completed), Swimming (1 km in 24 minutes)

SELECTED PUBLICATIONS

A = Accepted, S = Submitted SEE FULL LIST ON ▷ GOOGLE SCHOLAR

[1,A] **Chengjie Lu**, Shaukat Ali, and Tao Yue. (2024). EpiTESTER: Testing Autonomous Vehicles With Epigenetic Algorithm and Attention Mechanism. *IEEE Transactions on Software Engineering*, Vol. 50, Issue 10, pp. 2614-2632.

[2,A] **Chengjie Lu**, Yize Shi, Huihui Zhang, Man Zhang, Tiexin Wang, Tao Yue, and Shaukat Ali. (2023). Learning Configurations of Operating Environment of Autonomous Vehicles to Maximize their Collisions. *IEEE Transactions on Software Engineering*, Vol. 49, Issue 1, pp. 384-402.

[3,A] **Chengjie Lu**, Qinghua Xu, Tao Yue, Shaukat Ali, Thomas Schwitalla, and Jan F. Nygård. (2023). EvoCLINICAL: Evolving Cyber-Cyber Digital Twin with Active Transfer Learning for Automated Cancer Registry System. In *Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2023)*, pp. 19731984. ACM. 2023, San Francisco, CA, USA.

[4,A] **Chengjie Lu**, Jiahui Wu, Shaukat Ali, and Mikkel Labori Olsen. (2025). Assessing the Uncertainty and Robustness of the Laptop Refurbishing Software. In *2025 IEEE Conference on Software Testing, Verification and Validation (ICST)*, pp. 406-416. IEEE. 2025, Napoli, Italy.

[5,A] **Chengjie Lu**, Tao Yue, and Shaukat Ali. (2023). DeepScenario: An Open Driving Scenario Dataset for Autonomous Driving System Testing. In *2023 IEEE/ACM 20th International Conference on Mining Software Repositories (MSR)*, pp. 52-56. IEEE. 2023, Melbourne, Australia.

[6,A] Jiahui Wu, **Chengjie Lu**, Aitor Arrieta, Shaukat Ali, Thomas Peyrucain. (2025). Vision Language Model-based Testing of Industrial Autonomous Mobile Robots. In *2026 IEEE/ACM 48th International Conference on Software Engineering, Software Engineering in Practice (SEIP) Track*.

[7,A] Jiahui Wu, **Chengjie Lu**, Aitor Arrieta, Tao Yue, and Shaukat Ali. (2024). Reality Bites: Assessing the realism of driving scenarios with large language models. In *Proceedings of the 2024 IEEE/ACM First International Conference on AI Foundation Models and Software Engineering (FORGE '24)*, pp. 4051. ACM. 2024, Lisbon, Portugal.

[8,A] **Chengjie Lu**, Huihui Zhang, Tao Yue, and Shaukat Ali. (2021). Search-Based Selection and Prioritization of Test Scenarios for Autonomous Driving Systems. In *13th International Symposium on Search Based Software Engineering (SSBSE 2021)*, pp. 41-55. Springer. 2021, Bari, Italy.

[9,A] Yize Shi, **Chengjie Lu**, Man Zhang, Huihui Zhang, Tao Yue, and Shaukat Ali. (2021). Restricted Natural Language and Model-based Adaptive Test Generation for Autonomous Driving. In *2021 ACM/IEEE 24th International Conference on Model Driven Engineering Languages and Systems (MODELS)*, pp. 101-111. IEEE. 2021, Fukuoka, Japan.

[10,A] **Chengjie Lu**. (2023). Test Scenario Generation for Autonomous Driving Systems with Reinforcement Learning. In *2023 IEEE/ACM 45th International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)*, pp. 317-319. IEEE. 2023, Melbourne, Australia. **1st Place Winner at SRC@ICSE2023**.

[11,S] **Chengjie Lu**, Pablo Valle, Jiahui Wu, Erblin Isaku, Hassan Sartaj, Aitor Arrieta, and Shaukat Ali. Foundation Models for Software Engineering of Cyber-Physical Systems: the Road Ahead. Manuscript under major revision in *ACM Transactions on Software Engineering and Methodology (TOSEM)*.

[12,S] **Chengjie Lu**, Tao Yue, Man Zhang, and Shaukat Ali. Multi-Modal Reinforcement Learning-based Testing of Autonomous Driving Systems. Manuscript submitted for publication in *Journal of Systems and Software (JSS)*.

[13,S] **Chengjie Lu**, Jiahui Wu, Shaukat Ali, Malaika Din Hashmi, et al. UAMTERS: Uncertainty-Aware Mutation Analysis for DL-enabled Robotic Software. Manuscript submitted for publication in *ACM Transactions on Software Engineering and Methodology (TOSEM)*.

[14,S] Pablo Valle, **Chengjie Lu**, Shaukat Ali, and Aitor Arrieta. Evaluating Uncertainty and Quality of Visual Language Action-enabled Robots. Manuscript submitted for publication in *IEEE Transactions on Software Engineering*.