# Liang Zhang

**Mobile**: +86 13416661681 — **Email**: 13416661681@163.com
Google Scholar, Linkedin, GitHub

## Summary

Ph.D. from QS100 University and Senior Research Engineer with extensive experience in **RL**-assisted **(UAV)** communication network, **FL**-based traffic prediction, **distributed training in multiple zones for LLM**, and **LLM inference acceleration**. Having **end-to-end** capabilities ranging from demand mining, algorithm research, prototype verification, to POC support, as well as **management experience in university-industry cooperation projects**. Reviewer of **4+ CCF A** journals. Passionate about building scalable multi-agent systems and bridging academic research with production-grade deployment.

## Working Experiences

- **Saudi Aramco** Dharan, Saudi Arabia
  *Project Management Engineer* *2024 - 2025*
  - **AI assisted simulator development**: Designed GUI + AI-assisted history matching for the GigaPOWERS simulator
- **Huawei Technologies Co., Ltd.** Beijing/Shenzhen, China
  *Senior Research Engineer* *2022 - 2024*
  - **Cross-Zone Distributed LLM Training**:
    - Cooperate with the cross-functional team (Solutions, Marketing, Standards) to architect networks for **multi-domain**, **federated**, **compute-storage separation** and **train-infer separation** scenarios; invented **WAN lossless scheduling** & **key-frame identification** techniques, filed patents.
    - With **2012 Lab** & **super-computing centers**, prototyped **175 B-param model** training over **800 G** cross-domain links via **DeepSpeed**: $\leq 7\,\%$ latency increase and $\leq 0.5\,\%$ accuracy loss.
    - Co-owner of **Shanghai Telecom's 800 G distributed LLM kick-off**; maintained partnerships with **Xiamen Univ.**, **Jinan Super-computing Center** and **Zhejiang Univ.**; onboarded **Shandong Univ.**
  - **Video QoS Assurance**:
    - Deployed AI accelerator cards at edge gateways to extract 11-dim network KPIs (jitter, latency, etc.) in real time and stream them to cloud controller every 4 ms.
    - Designed bidirectional LSTM model with pruning that shrank search space by **35 %**, delivering **>95 %** accuracy in predicting multi-frame freezes within **5 s**.
    - Built in-house **Java LSTM-inference micro-service** (co-dev with Platform Team) sustaining **1 k** concurrent requests; integrated with **NCE** to raise alarms **8 s** earlier on average.
    - Delivered **3+** PoCs for **Guangdong Government** and enterprise clients, collected **100+ h** of traces and cut complaint rate by **42 %**.
  - **LLM Inference Acceleration & Lingqu AI Cluster Design**:
    - Organized discussion about **attention** & **KV-Cache optimization**, covering **vLLM**, **TensorRT**, **SGLang**, etc.
    - Implemented **Ascend-based flash/radix attention** optimizations cutting **first-token latency by 20 %** and **end-to-end latency by 30 %** in code-completion and CoT tasks; throughput improved significantly.
    - Presented results to internal (Process-IT, Equipment Dept., AI-Software Co-dev WG) and external customers (Meituan).
- **Saudi Company for Artificial Intelligence (SCAI)** Remote
  *R&D Consultant, Intern/Part time* *2021 - 2022*
  - **Delivered AI Strategy Workshop Series**: design and execution of eight in-depth AI-themed workshops for SCAI, covering education, healthcare, and smart-city domains, engaging cross-departmental stakeholders such as the Saudi Data and AI Authority (SDAIA), NEOM, and SenseTime.
  - **Cross-domain AI Scenario Implementation**: Aligned with Saudi Vision 2030, produced flagship use cases that integrate AI into education (adaptive-learning systems, NLP-based teaching tools), healthcare (early screening, precision patient tracking), and smart cities (traffic forecasting, 5G/6G convergence).

- **Knowledge Transfer & Localization**: epare state-of-art AI technology-selection guides—e.g., CNN for medical imaging and GANs for smart-city data augmentation—based on China and Middle-East experiences, and deliver the tutorial for local teams on Chinese open-source frameworks such as PaddlePaddle to lower technical barriers.

- **Saudi Aramco**
  *Deep Learning Engineer, Intern* *Jun. 2020 - Aug. 2020*
  - CNN-based history matching for upstream petroleum simulation.
- **NEOM**
  *Software Development Engineer, Intern* *Apr. 2019 - Jun. 2019*
  - IOS App development in the domain of food chains: Smartdinner.
- **Saudi Basic Industries Corporation (SABIC)**
  *Research Engineer, Intern* *Jun. 2018 - Aug. 2018*
  - Semiconductor device component analysis by using X-Ray diffraction.
- **Beijing Miyoshi Interactive Educational Technology Co., Ltd.**
  *Senior lecture, Part-time* *Sep. 2016 - Jun. 2017*
  - Delivering lectures of physics for middle school students (more than 500 students).

## Education

- **King Abdullah University of Science and Technology (KAUST)** Thuwal, Saudi Arabia
  *Doctor of Philosophy in Electrical and Computer Engineering* *Aug. 2018 - Jun. 2022*
  - **Supervisor**: Prof. Basem Shihada
  - **Thesis title**: Learning-Based Approaches for Next-Generation Intelligent Networks
- **King Abdullah University of Science and Technology (KAUST)** Thuwal, Saudi Arabia
  *Master of Science in Electrical Engineering* *Aug. 2017 - Aug. 2018*
  - **Supervisor**: Prof. Kazuhiro Ohkawa
- **University of Science and Technology Beijing (USTB)** Beijing, China
  *Bachelor of Science in Applied Physics* *Sep. 2012 - Jun. 2016*

## Highlighted Publications

- **L. Zhang**, A. Celik, S. Dang and B. Shihada, "Energy-Efficient Trajectory Optimization for UAV-Assisted IoT Networks" in *IEEE Transactions on Mobile Computing*, vol. 21, no. 12, pp. 4323-4337, 1 Dec. 2022 (one of the fifty most popular papers of TMC).
- **L. Zhang**, C. Zhang and B. Shihada, "Efficient Wireless Traffic Prediction at the Edge: A Federated Meta-Learning Approach" in *IEEE Communications Letters*, vol. 26, no. 7, pp. 1573-1577, Jul. 2022.
- **L. Zhang**, C. Zhang, S. Dang and B. Shihada, "Lessons from the Commercial Failure of Project Loon for 6G Research Roadmap Design" in *Frontiers in Communications and Networks*, vol. 3, 2022.
- **L. Zhang**, W. Abderrahim and B. Shihada, "Heterogeneous Traffic Offloading in Space-Air-Ground Integrated Networks" in *IEEE Access*, vol. 9, pp. 165462-165475, Dec. 2021.
- **L. Zhang**, G. Ma, A. Al-Ghadhban, S. Dang and B. Shihada, "ICAQ: Adaptive QoS System for 5G and Beyond Applications" in *IEEE ICCT*, Nanning, China, 2020.
- L. Bai, **L. Zhang***, G. Zhang, L. Zhang, P Medagliani, and S. Martin, "A Distributed Congestion Mitigation Mechanism Based on Neighboring Nodes Traffic Steering' *International Conference on Network of the Future* , Izmir, Turkey, 2023.
- X. Rao, H. Wang, **L. Zhang**, J. Li, S. Shang, P. Han, "FOGS: First-Order Gradient Supervision with Learning-based Graph for Traffic Flow Forecasting" in *IJCAI*, Messe Wien, Vienna, Austria, 2022
- Y. Xie, W. Pei, D. Guo, **L. Zhang**, H. Zhang, X. Guo, X. Xing, X. Yang, F. Wang, Q. Gui, Y. Wang, H. Chen, "Improving adhesion strength between layers of an implantable parylene-C electrode" in *Sensors and Actuators A: Physical*, vol. 260, pp. 117-123, 2017.
- F. Wang, D. Guo, Y. Xie, **L. Zhang**, W. Pei, H. Chen, "An implantable optrode composed of fiber and flexible thin-film electrode" in *Optoelectronics Letters*, vol. 14, no. 4, pp. 271-275, 2018.
- C. Zhang, G. Ma, **L. Zhang**, and B. Shihada, "Graph Convolutional Networks Empowered Origin-Destination Learning for Urban Traffic Prediction" *CAAI Transactions on Intelligence Technology.*
- International patent WO2025067200A1:DATA PROCESSING METHOD AND RELATED DEVICE
- Chinese patent CN119449662A: Time delay measurement methods, devices, equipment and computer-readable storage media

## Professional Reviewing Experiences

- **Reviewer**: IEEE TMC, ComMag, TWC, TCOM, TVT, WCL, CL, SensorJ, IoTJ, Access, GlobeCom, Scientific Reports

- **Speaking**: Keynote Speaker – V-Electrical 2023; TPC Member – IEEE VTC Fall 2022

## Grants

- **KAUST**: PhD Fellowship (2018–2022), MS Scholarship (2017–2018)
- **China**: National Scholarship (MOE), Huang Kun Prize (CAS)

## Technical and Soft Skills

- **Languages**: Python, Java, Shell, Swift, MATLAB, C++, LaTeX
- **Frameworks**: Gym, Torch, TensorFlow, Ryu, POX, Mujoco, vLLM, Deepspeed, TRT, vLLM, sglang, langchain
- **Algorithm**: [**RL**] Q-learning, Sara, A2C, A3C, DDPG, PPO, DPO, GRPO, DAPO; [**FL**] FedAvg, FedProx; [**Meta-learning**] MAML, Reptile
- **Systems**: Shaheen/Ibex, Linux, Apple OS, Data center & distributed clusters

## CERTIFICATIONS

**Coursera**: Deep Learning Specialization, TensorFlow, Developer Specialization, Preparing for Google Cloud Certification: Cloud Engineer Specialization, Machine Learning with TensorFlow on Google Cloud Platform Specialization