# Yuyang Du

Email: yuydu@ie.cuhk.edu.hk, duyuyang01@gmail.com | Google Scholar| Personal Homepage

## Education Background

**Ph.D, Information Engineering** 01/2021-06/2025
The Chinese University of Hong Kong

**B.Eng., Electronic Engineering** 09/2015-07/2019
Peking University

## Professional Experiences

**The Chinese University of Hong Kong, Postdoctoral Fellow** 07/2025-Present
Research: Generative AI and Wireless Communications for Mobile Applications (advisor: Prof. Soung Liew).

**University of California, San Diego, Visiting Postdoctoral Researcher (Scheduled)** 08/2026-07/2027
Research: AI-driven wireless system design focusing on 6G ISAC (host: Prof. Nuria Gonzalez Prelcic)

**Harvard University, Visiting PhD Student** 11/2024-05/2025
Research: LLM inference acceleration and reasoning (hosts: Prof. Minlan Yu and Prof. Marinka Zitnik)

**Huawei, 2012 Lab, Communication Engineer** 08/2019-12/2020
Work: Algorithm design for 5nm smartphone SoCs, focusing on PHY signal processing in 5G modems.

## Research Experiences (a list of publications available at the end of the CV)

I develop intelligent mobile systems, with applications in industrial automation and smart healthcare. My toolkit for building such systems includes efficient edged LLMs, task-oriented communication, network optimization, and robotics. My recent work addresses two challenges: 1) **cost-effective generative AI for resource-constrained edge**, and 2) **reliable, task-oriented communications for real-time mobile applications**. I have published **27** papers in top-tier journals including IEEE TWC/TCOM/TMLCN/TVT/TMI/ComMag, and leading conferences in AI, networking, and robotics, including ICLR/NeurIPS/ICCV/Mobisys/ICRA, with **25** as a major author (i.e., first/co-first author, project lead, or corresponding author).

### 1. Cost-Effective AI at Edge (making LLMs powerful enough for edge applications despite limited resources)

- **Edge MoE Inference:** Multi-device MoE inference frameworks leveraging distributed IoTs for collaborative LLM inference. Through precise expert-activation prediction and dynamic expert offloading, we achieved 75% decoding throughput while using 1/3 GPU memory in total. Our framework reduces the per-node GPU memory requirement to <1GB, falling within a typical edge device's computational budget [A5, *under review* OSDI'26].
- **Edge Agents:** LLM agents tailored to IoT applications. IndusGCC we developed an LLM-supervised device control agent for industrial IoT automation, which could automatically operate large-scale machines and robotic arms via GUI control interfaces [C9-NeurIPS'25]. Cellular-X developed a voice-controlled agent for 5G/LTE network management. We implemented a real-time cellular system with software-defined radio and deployed the agent system for human interactions and network adjustments [C7-Mobisys'25].
- **LLM Post-Training:** Model post-training pipelins that enable compact LLMs to rival larger models on targeted tasks, making them viable for edge deployment. Methodologies include supervised finetuning [C6-ICNC'25] and model alignment [A2-Nature Communications].

2. **Wireless Communication at Edge (connecting distributed nodes via wireless networks tailored to IoT)**
   - **URLLC:** Prototype-driven research on ultra-reliable low-latency communication, with core algorithms validated on software-defined radio (SDR) platforms. Topics include signal detection [J7-IEEE TVT], multiple access [J5-IEEE TWC, A4], network optimization [J1-IEEE TCOM], and PA nonlinearity-aware power optimization for MIMO-OFDM systems [J4-IEEE TVT, J6-IEEE SJ].
   - **Task-Oriented Communication:** LLM/MLLM-powered semantic communication that prioritizes task-level meaning over bit accuracy, maintaining reliable task performance even when underlying channels are unstable. Research topic includes variable-length Joint Source-Channel Coding (JSCC) with context and task awareness [C10-ICNC'26], and anti-out-of-distribution (anti-OOD) problems in image JSCC [C10-ICNC'26, C5-Wiopt'25].
3. **AI-Wireless Co-design for Edge Applications (focusing on industrial automation and remote healthcare)**
   - **Industrial IoT (IIoT) Automation:** IIoT management and control framework powered by LLMs [J9-IEEE ComMag, A3], particularly in AI-powered production line coordination and wireless-connected robotic collaboration for smart manufacturing, with an ongoing pilot in pharmaceutical production [C8-ICCV, A1].
   - **Remote Healthcare:** AI-augmented clinical robots and medical agents for 5G/6G-enabled remote healthcare [C12-ICLR'26, C4-ICRA'24, J8-IEEE TMI].

## Funding Experiences (further participation details available at the link)

Since 2023, I have had the opportunity to contribute to the preparation and execution of funding projects within my group, primarily serving as the **major proposal writer** and participating in a total of **~32.7 million HKD** applications. Among these, **~11.4 million HKD** are currently approved or pending for results. I have also gained hands-on experience in preparing large-scale collaborative proposals for CRF-2024 and CRF-2025, in collaboration with PIs from CUHK, HKUST, HKU, Imperial College London, Aalborg University, and Singapore University of Technology and Design.

1. **UGC General Research Fund (~1.4m HKD)** **2026-2029 (Pending)**
   - **Topic:** Distributed MoE Inference at the 6G Edge: Ultra-Accurate Expert-Activation Prediction, Partial Expert Residency, and Dynamic Token Handling for Batched Systems
   - **Role:** Co-Investigator, contributed to drafting the full proposal.
2. **ITC Mainland-Hong Kong Technology Cooperation Funding Scheme (~2m HKD)** **2026-2028 (Pending)**
   - **Topic:** SmartLink Industry 4.0: A Trusted Collaboration System for Industry 4.0 Integrating Large Models, Ultra-Reliable Low-Latency Networks, and Blockchain
   - **Role:** Co-Investigator, led the drafting of the full proposal and coordination with industry sponsors.
3. **ITC Mainland-Hong Kong Technology Cooperation Funding Scheme (~3.2m HKD)** **2026-2028 (Pending)**
   - **Topic:** Research on High-Reliability Transmission and Intelligent Assurance Technologies for Remote Surgery
   - **Role:** Co-Investigator, led the drafting of the full proposal and coordination with industry sponsors.
4. **ITC Innovation and Technology Support Program (~1.4m HKD)** **2026-2028 (Conditionally Approved)**
   - **Topic:** Multimodal LLM-Driven Semantic Communication for Vehicular Networks
   - **Role:** Co-Investigator, led the drafting of the full proposal.
5. **STIC SZ-HK Science and Technology Collaboration Fund (~3.4m HKD)** **2024-2027 (Approved)**
   - **Topic:** Multi-stream Ultra Reliable Networking and Just-in-Time Communications for High Performance IoT
   - **Role:** Co-Investigator, contributed to drafting the HK-side proposal and led the project executions.
6. **UGC Funding Scheme for Innovative Technology in Education (~0.3m HKD)** **2024-2025 (Completed)**

- **Topic:** Academia-Industry Collaboration for LLM Education: Hands-on Experiences and Co-Training Opportunities
- **Role:** Key Member, drafted the proposal, led the project execution, and wrote the project completion report.

**7. CUHK Research Committee Direct Grant (~50k HKD)** **2024-2026 (Completed)**

- **Topic:** Robust Packet Detection for Short-packet Communications
- **Role:** Key Member, drafted the proposal, led the project execution, and wrote the project completion report.

## Mentorship

I have had the honor to work with talented students and research assistants from the *Institute of Network Coding* at IE Department, CUHK (led by Prof. Soung Chang Liew) and *Institute of Medical Intelligence* at CSE Department, CUHK (led by Prof. Pheng Ann Heng). A list of my mentees is as follows:

- **PhD Students (10)**
  - Liujianfu Wang (IE-24Fall) – ACM Mobisys'25, ICNC'25, TMLCN'26, OSDI'26 (*submitted to*)
  - Feifan Zhang (IE-24Fall) – Wiopt'24, TMLCN'25, ICNC'26, TWC'26 (*submitted to*).
  - Shiqi Xu (IE-24Fall) – Wiopt'26, IoTJ'26 (*submitted to*)
  - Yirun Wang (IE-25Fall) – IEEE GLOBECOM'26 (*submitted to*)
  - Yuchan Pan (IE-25Fall) – ACM Mobisys Demo'26 (*submitted to*)
  - Qun Yang (IE-20Fall, collaboration since 2024) – IEEE JSAC (*submitted to*)
  - Hongwei Cui (IE-21Fall, collaboration since 2024) – IEEE Communications Magazine'25
  - Kexin Chen (CSE-22Fall, collaboration since 2024) – ICRA'24, TMI'25, ICCV'25
  - Yiyi Zhang (CSE-24Fall, collaboration since 2024) – Nature Communications'26 (*major revision*)
  - Xinlin Dang (CSE-23Fall, collaboration since 2025) – ICLR'26
- **Master Students (9)**: Wenyuan Han (2026), Zixuan Wei (2026), Yunpeng Yang (2025), Zhicong Zhang(2025), Yifan Xiang (2025), Xiaoran Yang (2025), Xiaoyan Liu (2025), Xinyi Long (2025), Hongxu Chen (2024)
- **Research Assistants (7)**: Yue Zhan, Yining Du, Haomin Qi, Jingqi Lin, Lihao Zhang, Qiying Hu, Xingyu Chen
- **CUHK Final Year Project (6)**: Mingyue Zhang (2025), Ziheng Kang (2025), Lik Hang Chan (2024), Yufan Huang (2024), WinHei Leung (2024), Chim Long You (2023)

## Research Talk

- Cellular-X: An LLM-empowered Cellular Agent for Efficient Base Station Operations 2025
  *Presented and Demonstrated at ACM Mobisys 2025 (Anaheim, CA, USA)*
- RaC: Fine-Tuning LM for Enhanced Understanding of Communication and Computer Networks 2025
  *Presented at IEEE ICNC 2025 (Honolulu, HI, USA)*
- Addressing Out-of-Distribution Challenges in Image Semantic Communication Systems with MLLMs 2024
  *Presented at IEEE WiOpt 2024 (Seoul, South Korea)*
- LLM for Complex Signal Processing in FPGA-based Software Defined Radios: A Case Study on FFT 2024
  *Presented at IEEE VTC 2024 (Online)*
- SER Optimization for Nonlinear MIMO-OFDM Systems (PKU EECS Undergraduate Visiting Program) 2018
  *University of Limerick, hosted by Prof. Yiming Lei and Prof. Sean McGrath (Limerick, Ireland)*

## Award & Honors

- Reach Out Award, Graduate School, CUHK 2025

- Rising Star Award (1%), HiSilicon, Huawei 2020
- Outstanding Undergraduate Thesis (top 5%), Department of Electronic Engineering, Peking University 2019
- National Scholarship (top 5%), Peking University 2018

## Academic Service

- Journal Reviewer: *IEEE TCOM / TWC / TSP / TII / TVT / THMS / TNNLS / TMM / ComMag / WCL / CL / Access, ACM TIoT / TKDD, Nature Communications, Scientific Reports*
- Conference Reviewer: *IEEE ICC, IEEE GLOBECOM, IEEE ICCC, IEEE WCNC, AAAI, NeurIPS, ICRA*

## Teaching Experience

- Undergraduate Teaching Assistant, Digital Circuit (04834610), PKU, 2018 Fall
- Graduate Teaching Assistant, Linear Algebra (ENGG1120), CUHK, 2021 Spring / 2022 Spring
- Graduate Teaching Assistant, Computer Networks (IERG3310), CUHK, 2021 Fall /2023 Fall
- Graduate Teaching Assistant, Basic Analog and Digital circuits (IERG2060), CUHK, 2024 Spring
- Laboratory Tutor, Electronic Circuit Design Laboratory (IERG1810), CUHK, 2022 Fall/2023 Spring
- Laboratory Tutor, Information and Software Engineering Practice (IERG3080), CUHK, 2024 Fall

## Publications (* for equal contribution, # for corresponding author/project lead, underline for students co-advised)

**Journal**

- [J10] Zhang, F.*, **Du, Y.***, Chen, K., Shao, Y., Liew, S. C., "Out-of-Distribution in Image Semantic Communication: A Solution with Multimodal Large Language Models." *IEEE Transactions on Machine Learning in Communications and Networking*, 2025.
- [J9] Cui, H.*, **Du, Y.***, Yang, Q., Shao, Y., Liew, S. C., "LLMind: Orchestrating AI and IoT with LLMs for complex task execution." *IEEE Communications Magazine*, 2025.
- [J8] **Du, Y.**, Chen, K., Zhan, Y., Low, C.H., Islam, M., Guo, Z., Jin, Y., Chen, G., Heng, P. A. " LMT++: Adaptively Collaborating LLMs with Multi-specialized Teachers for Continual VQA in Robotic Surgical Videos." *IEEE Transactions on Medical Imaging*, 2025.
- [J7] **Du, Y.**, Liew, S. C., "Reliable Packet Detection for Random Access Networks: Analysis, Benchmark, and Optimization." *IEEE Transactions on Vehicular Technology*, 2025.
- [J6] **Du, Y.**, Hao, L., Lei, Y., "Nonlinear Multi-Carrier System with Signal Clipping: Measurement, Analysis, and Optimization." *IEEE Systems Journal*, 2024.
- [J5] **Du, Y.**, Liew, S. C., Shao, Y., "Efficient FFT Computation in IFDMA transceivers." *IEEE Transactions on Wireless Communications*, 2023.
- [J4] **Du, Y.**, Hao, L., Lei, Y., "SER Optimization in OFDM-IM Systems with Nonlinear Power Amplifiers." *IEEE Transactions on Vehicular Technology*, 2023.
- [J3] **Du, Y.**, Lei, Y., McGrath, S., "SER optimization in transparent OFDM relay systems in the presence of dual nonlinearity." *Digital Signal Processing*, 2022.
- [J2] **Du, Y.**, Chen, J., Lei, Y., Hao, X., "Performance analysis of nonlinear spatial modulation multiple-input multiple-output systems." *Digital Signal Processing*, 2021.
- [J1] Shao, Y., Liew, S. C., Chen, H., **Du, Y.**, "Flow sampling: Network monitoring in large-scale software-defined IoT networks." *IEEE Transactions on Communications*, 2021.

**Conference**

- [C12] Dang, X., Chen, K., Su, X., …, Long X., **Du, Y.**, Zitnik, M., Heng, P. A." KnowGuard: Knowledge-Driven Abstention for Multi-Round Clinical Reasoning", *ICLR 2026*
- [C11] Qi, H.***, Du, Y.***, Zhang, L., Liew, S. C., Chen, K., Du, Y. "VeriRAG: A Retrieval-Augmented Framework for Automated RTL Testability Repair", *IEEE ISQED 2026*.

- [C10] Zhang, F.*, **Du, Y.***, Xiang, Y., Liu, X., Liew, S. C. " SA-OOSC: A Multimodal LLM-Distilled Semantic Communication Framework Leveraging Contextual Understanding for Coding Efficacy", *IEEE ICNC 2026*.
- [C9] Yang, X., **Du, Y.#**, Chen, K.#, Liew, S. C.#, …, Heng, P. A.# "IndusGCC: A Data Benchmark and Evaluation Framework for GUI-Based General Computer Control in Industrial Automation." *NeurIPS 2025*.
- [C8] Chen, K., **Du, Y.#**, Li, J., Cao, H., Guo, M., Dang, X., Li, L., Qiu, J., Chen, G.#, Heng, P. A. "ChemMiner: A Large Language Model Agent System for Chemical Data Mining." *ICCV* 2025.
- [C7] Wang, L.*, **Du, Y.***, Long, X., Liu, X., Chen, K., Liew, S. C., " Cellular-X: An LLM-empowered Cellular Agent for Efficient Base Station Operations." *ACM Mobisys 2025*.
- [C6] Wang, L.*, **Du, Y.***, Lin, J., Chen, K., Liew, S. C. "Rephrase and Contrast: Fine-Tuning Language Models for Enhanced Understanding of Communication and Computer Networks." *IEEE ICNC 2025*.
- [C5] Zhang, F.*, **Du, Y.***, Chen, K., Shao, Y., Liew, S. C., "Addressing Out-of-Distribution Challenges in Image Semantic Communication Systems with Multi-modal Large Language Models." *IEEE WiOpt 2024*.
- [C4] Chen, K.*, **Du, Y.***, You, T., Islam, M., Guo, Z., Jin, Y., Chen, G., Heng, P. A. "LLM-Assisted Multi-Teacher Continual Learning for Visual Question Answering in Robotic Surgery." *IEEE ICRA 2024*.
- [C3] **Du, Y.**, Deng, H., Liew, S. C., Chen, K., Shao, Y., Chen, H., " LLM for Complex Signal Processing in FPGA-based Software Defined Radios: A Case Study on FFT." *IEEE VTC 2024*.
- [C2] **Du, Y.**, Hao, L., Lei, Y., "SER Analysis and Joint Optimization in Nonlinear MIMO-OFDM Systems with Clipping." *IEEE VTC*, 2023.
- [C1] **Du, Y.**, Hao, L., Liu, Z., Chen, Y., Lei, Y., "Ergodic Rate Performance in Nonlinear Omnidirectional Coding MIMO-OFDM Systems." *IEEE UEMCON*, 2019.

**Arxiv Preprint**

- [A5] Wang, L.*, **Du, Y.*#**, Pan, Y., Liew, S. C.#, Liu, Y., Chen, K. "OD-MoE: On-Demand Expert Loading for Cacheless Edge-Distributed MoE Inference", https://arxiv.org/abs/2512.03927.
- [A4] Xu, S.*, Zhang, L.*, **Du, Y.***, Liew, S. C. "A Hybrid TDMA/CSMA Protocol for Time-Sensitive Traffic in Robot Applications", Available: https://arxiv.org/abs/2509.06119.
- [A3] **Du, Y.**, Yang, Q., Wang, L., Lin, J., Cui, H., Liew, S. C. "LLMind 2.0: Distributed IoT Automation with Natural Language M2M Communication and Lightweight LLM Agents", https://arxiv.org/abs/2508.13920.
- [A2] Zhang, Y., Chen, X., Chen, K., **Du, Y.#**, Dang, X., …, Heng, P. A., "The Dual-use Dilemma in LLMs: Do Empowering Ethical Capacities Make a Degraded Utility?." https://arxiv.org/abs/2501.13952.
- [A1] Chen, K., Li, J., Wang, K., **Du, Y.#**, …, Heng, P. A., "Chemist-X: Large Language Model-empowered Agent for Reaction Condition Recommendation in Chemical Synthesis.", https://arxiv.org/abs/2311.10776.

**Patent**

- [P2] Liew, S. C., Zhang, F., **Du, Y.**, Chen, K. "An Image Semantic Communication Method Based on Multimodal Large Language Models", CN Application No: 202411403556.5
- [P1] Wang, L., **Du, Y.**, Chen, K., Liew, S. C." A Fine-Tuning Method for Networking Large Language Models", CN Application No: 202510011166.1

## References


**Prof. Soung Chang Liew**
Choh-Ming Li Professor in Information Engineering
The Chinese University of Hong Kong
soung@ie.cuhk.edu.hk
*Relationship: PhD & Postdoctoral Advisor*

**Prof. Pheng Ann Heng**
Choh-Ming Li Professor in Computer Science and Engineering
The Chinese University of Hong Kong
pheng@cse.cuhk.edu.hk
*Relationship: Academic Mentor*

**Prof. Yulin Shao**
Assistant Professor in Electrical and Electronic Engineering
University of Hong Kong
ylshao@hku.edu.hk
*Relationship: Academic Mentor*

**Prof. Yiming Lei**
Associate Professor in Electronic Engineering
Peking University
leiym@pku.edu.cn
*Relationship: Academic Mentor*