

## ABOUT

I am an M.Sc. student at **Peking University** (advised by Prof. Wenjia Zhang & Prof. Guanhua Chen), with a B.E. from **SUSTech**, currently a Research Intern at **Microsoft Research Asia** (GenAI Group, Dr. Li Dong).

My research centers on **post-training data for LLMs** what makes it good, how to synthesize and select it at scale (**FANNO**, **InstructDiff**, **Tag-Instruct**, **AlignDiff**), and how to design training objectives that acquire capabilities efficiently without forgetting what the model already knows (**ASFT**, **CFT**). I have also applied this perspective to domain-specific foundation models for urban intelligence (**PlanGPT**, **PlanGPT-VL**, **UrbanClaw**).

## EDUCATION

- 🏛️ **Peking University** Beijing, China  
*M.Sc. in Smart City & Data Sciences* 2024 – Present
- **Advisors:** Prof. Wenjia Zhang, Prof. Guanhua Chen.
  - **Research:** Data-centric LLM post-training, on-policy distillation, alignment.
- 🏛️ **Southern University of Science and Technology (SUSTech)** Shenzhen, China  
*B.E. in Computer Science* 2020 – 2024
- **Advisors:** Prof. Zipei Fan, Prof. Xuan Song.
  - GPA 90.2/100 (Top 10%); Outstanding Graduate & Outstanding Thesis (Top 5%, 2024).
  - **Visiting:** 🌸 **UTokyo CSIS** (2022, 2023); 🏛️ **NUS CS** (2022).

## EXPERIENCE

- Microsoft Research Asia · GenAI Group** Beijing, China  
*Research Intern*, supervised by Dr. Li Dong. Nov 2025 – Present  
On-policy distillation and agent harness.
- Shanghai AI Laboratory · OpenData Lab** Shanghai, China  
*Research Intern*, foundation language models. May 2025 – Oct 2025
- SenseTime · Foundation Language Model Center** Shenzhen, China  
*Research Intern*, foundation models. Jun 2024 – Sep 2024
- LocationMind** Tokyo, Japan  
*Project Assistant*, autonomous driving algorithms. Jun 2023 – Dec 2023

## PROJECTS

- PlanGPT Series** [plangpt.github.io](https://plangpt.github.io)  
First systematic study of LLMs for urban planning. Includes **PlanGPT** (text LLM for scheme generation), **PlanGPT-VL** (VLM for planning maps), **PlanGPT-R1** (reasoning-enhanced model), and **UP-Bench / PlanBench-V** (benchmarks).
- UrbanClaw** [app.urbanclaw.net](https://app.urbanclaw.net)  
AI-powered urban planning agent with multi-agent collaboration, Docker sandbox, MCP client layer, and self-evolving skill system.

## PUBLICATIONS

---

*First Author / Corresponding Author*

**PlanGPT: Enhancing Urban Planning with a Tailored Language Model** urban AI

- He Zhu, Guanhua Chen, Wenjia Zhang<sup>†</sup>
- *ACL Industry 2025* · **Oral** [paper]

**Anchored Supervised Fine-Tuning** alignment

- He Zhu<sup>\*</sup>, Junyou Su<sup>\*</sup>, Peng Lai<sup>\*</sup>, Wenjia Zhang, Linyi Yang, Guanhua Chen<sup>†</sup>
- *ICLR 2026* [paper]

**InstructDiff: Domain-Adaptive Data Selection via Differential Entropy** data-centric

- Junyou Su<sup>\*</sup>, He Zhu<sup>\*†</sup>, Guanhua Chen<sup>†</sup>
- *ACL 2026 (Main)* [paper]

**FANNO: Augmenting High-Quality Instruction Data with Open-Sourced LLMs Only** data-centric

- He Zhu, Yifan Ding, Yicheng Tao, ..., Guanhua Chen<sup>†</sup>
- *ACL 2025 (Findings)* [paper]

**Tag-Instruct: Controlled Instruction Complexity Enhancement** data-centric

- He Zhu, Zhiwen Ruan, Junyou Su, ..., Guanhua Chen<sup>†</sup>
- *ACL 2025 (Findings)* [paper]

**PlanGPT-VL: Vision-Language Model for Urban Planning Maps** urban AI

- He Zhu<sup>\*</sup>, Junyou Su<sup>\*</sup>, Minxin Chen<sup>\*</sup>, ..., Wenjia Zhang<sup>†</sup>
- *EMNLP Industry 2025* [paper]

**Personalized Individual Trajectory Prediction via Meta-Learning** urban AI

- He Zhu, Liyu Zhang, Zipei Fan
- *SIGSPATIAL 2022* · **Oral** [paper]

**AlignDiff: Exploiting Model-Intrinsic Information for Better Data Selection** data-centric

- Peng Lai<sup>\*</sup>, He Zhu<sup>\*</sup>, Zhiwen Ruan, ..., Yang Liu, Guanhua Chen<sup>†</sup>
- *Under Review, 2026* [paper]

**Towards Fair and Comprehensive Evaluation of Routers in Collaborative LLM Systems** efficiency

- Wanxing Wu<sup>\*</sup>, He Zhu<sup>\*</sup>, Yixia Li<sup>\*</sup>, Yun Chen, Guanhua Chen<sup>†</sup>
- *Under Review, 2026* [paper]

*Selected Co-authored Papers*

**Enhancing LLM Reasoning via Selective Critical Token Fine-Tuning** alignment

- *Under Review, 2026* [paper]

**Dripper: Token-Efficient Main HTML Extraction with a Lightweight LM** alignment

- *Under Review, 2026* [paper]

**Topic Over Source: The Key to Effective Data Mixing for LLM Pre-training** data-centric

- *Under Review, 2026* [paper]

**LayAlign: Enhancing Multilingual Reasoning via Layer-wise Adaptive Fusion** alignment

- *NAACL 2025 (Findings)*

**ToolExpNet: Optimizing Multi-Tool Selection in LLMs with Experience Networks** efficiency

- *ACL 2025 (Findings)*

**HHGNN: Heterogeneous Hypergraph Neural Network for Trajectory Prediction** urban AI

- *ICRA 2024*

## HONORS & SERVICE

---

- Outstanding Youth League Secretary, Peking University (**Top 5%**, 2025).
- Outstanding Graduate & Outstanding Thesis, Dept. of CS, SUSTech (**Top 5%**, 2024).
- Annual Outstanding Student, SUSTech (2021, 2022, 2023).
- **Reviewer:** ACL Rolling Review (ARR).