

Zhehao Dong

(+86) 13858551943 • dongzhehao@stu.pku.edu.cn
Research Interests: Embodied AI, World Model



EDUCATION

Peking University	Mechanical Engineering	Master	Sept. 2024 ~ Jun. 2027
• GPA: 85.60 / 100	• Rank: 4 / 79		
Sichuan University	Software Engineering & Mechanics	Bachelor	Sept. 2020 ~ Jun. 2024
• GPA: 90.98 / 100	• Rank: 1 / 23		
Major: Operating Systems (92), Data Structures and Algorithms (90), Discrete Mathematics (93), Digital Logic (91), Database Systems and Information Management (93)			
Awards: National Scholarship (Dec. 2021, Dec. 2023), Outstanding Graduate of Sichuan Province (Mar. 2024)			

PUBLICATIONS

- [1] EMMA: Enhancing Robotic Control Robustness through a Generative Visual Model
Zhehao Dong, Xiaofeng Wang, Zheng Zhu, Other Authors (**First Author, CoRL 2026, Accepted**)
- [2] GigaBrain-0: A World Model-Powered Vision-Language-Action Model
Zhehao Dong, Other Authors (Alphabetic Order)
Technical Report, 2025, **2.6k**
- [3] GigaWorld-0: World Models as Data Engine to Empower Embodied AI
Zhehao Dong, Other Authors (Alphabetic Order)
Technical Report, 2025, **1.6k**
- [4] HumanoidSpatial: Benchmarking Egocentric Multi-View Spatial Perception and Reasoning for Humanoid Robots in the Wild (**Co-First Author, Submitted to CVPR 2027**)
- [5] CFD-copilot: Leveraging Domain-adapted Large Language Model and Model Context Protocol to Enhance Simulation Automation
Zhehao Dong, Shanghai Du, Zhen Lu, Yue Yang
Chinese Journal of Aeronautics, 2026 (CAS Q1 Top Journal, Selected as **Cover Article**)
- [6] Fine-tuning a Large Language Model for Automating Computational Fluid Dynamics Simulations
Zhehao Dong, Zhen Lu, Yue Yang
Theoretical and Applied Mechanics Letters, 2025, **Citations: 70**

INTERNSHIP EXPERIENCE

GigaAI (Valuation > 20B CNY) Beijing Jul. 2025 ~ Jun. 2026
Embodied Foundation Model Group Research Intern *Mentor: Dr. Zheng Zhu, 23000+ Google Citations*

Project 1: Generalizable VLA Manipulation via Multi-View Embodied World Models

- **Overview:** Core contributor to the development of the embodied foundation models GigaWorld-0 and GigaBrain-0, responsible for world model architecture design and training, embodied data generation, VLA model training, and simulation/real-robot evaluation.
- **DiT-based Multi-View World Model Architecture Design:** Designed the *DreamTransfer* world model with a multi-view joint modeling mechanism and a two-stage low-/high-resolution training strategy, improving multi-view consistency by **42%** over SOTA methods. Integrated ControlNet to further enhance geometric consistency. *DreamTransfer* improved real-robot task success rate by **92%**.
- **Hard-Sample-Aware VLA Training with AdaMix:** Developed trajectory quality metrics and used policy evaluation results to dynamically schedule training data, yielding a further **17%** improvement in task success rate.
- **Real-Robot and Simulation Deployment:** Deployed VLA policy on the Agilex Cobot Magic platform for complex long-horizon tasks including clothes folding and tabletop cleanup, and conducted systematic generalization evaluations in RoboTwin and RoboLab simulation environments.
- **Paper:** *EMMA: Enhancing Robotic Control Robustness through a Generative Visual Model*. (CoRL, First Author)

Project 2: Egocentric Multi-View Spatial Perception and Reasoning Benchmark for Humanoid Robots

- **Overview:** Proposed the HSR-Bench to address the limitations of existing spatial intelligence benchmarks. Built a six-camera panoramic data acquisition platform aligned with the perception height and camera configuration of humanoid robots, covering 11 categories of real indoor and outdoor environments and producing 2,075 high-quality spatial reasoning VQA.
- **Contributions:** 1. Designed a hierarchical multi-view spatial reasoning task taxonomy consisting of 25 question types organized along the dimensions of spatial scope and reasoning complexity. 2. Built a scalable data construction and evaluation pipeline combining task-constrained generation, rule-based filtering, and human verification.
- **Insights:** The results revealed that current embodied foundation models are mainly bottlenecked by cross-view alignment and spatial state updating after motion. Proposed enhancing cross-view attention, explicit motion modeling, and temporal state memory to improve global spatial modeling and state tracking during robot motion.
- **Paper:** *HumanoidSpatial: Benchmarking Egocentric Multi-View Spatial Perception and Reasoning for Humanoid Robots in the Wild*. (Submitted to CVPR 2027, Co-First Author)

Meituan Inc. 📍Beijing

Apr. 2025 ~ Jul. 2025

Food Delivery Main Site Algorithm Team LLM Algorithm Intern

- **Overview:** Led the development and production deployment of the query-generation feature for the AI chat box on Meituan Food Delivery's homepage. Personalized query suggestions using users' historical information, current behavioral trajectories, and real-time contextual signals, reducing user input effort and improving the convenience and proactiveness of AI-assisted interactions.
- **DAPO Reinforcement Learning:** Explored multiple RL alignment algorithms including GRPO/DAPO (on-policy) and DPO/KTO (off-policy), and designed a multi-dimensional rubric-Based reward. Designed fault-tolerant reward mechanisms to effectively mitigate gradient vanishing caused by sparse rewards. Achieved an additional 15.7% CTR improvement with multi-node multi-GPU distributed training on A100 clusters.

Baidu Inc. 📍Shanghai

Nov. 2023 ~ May 2024

Baidu AI Cloud Group (ACG) VLM Algorithm Intern

- Implemented CogVLM for document extraction tasks. Built multimodal datasets for document understanding, covering visual question answering and structured field extraction. Conducted distributed training on A100 GPUs using DeepSpeed with LoRA. Improved JSON field-level accuracy to 90% (4% higher than internal baseline) by expanding and balancing the dataset and freezing high/low-resolution ViTs.
- Optimized semantic matching based long paragraph extraction. Addressed information truncation caused by long texts and large label sets by decomposing long label sequences into subdomains. Designed attention masks to isolate label domains, ensuring each label is predicted independently based only on relevant text.
- Optimized short field extraction pipeline and containerized services. Modified data scripts to inject negative samples automatically, mitigating false positives from missing negatives during training. Collaborated with engineering teams to package training, offline inference, and online serving into Docker images.

🏠 RESEARCH EXPERIENCE

End-to-End CFD Simulation Framework Based on Multi-Agent Framework Oct. 2024 ~ Mar. 2025

Project page: <https://github.com/YYgroup/AutoCFD>

- **Key Innovations and Contributions:** 1. Constructed the first natural language-driven dataset for OpenFOAM simulations. 2. Designed a multi-agent framework to automate the workflow from natural language requirements to full simulation results. 3. Our method achieved a pass@1 rate of 82.6% in incompressible flow scenarios.
- **Technical Highlights:** 1. Dataset construction: Performed data augmentation from both configuration parameter and semantic expression perspectives; added Chain-of-Thought reasoning to outputs to enhance their logical consistency. 2. SFT: Conducted LoRA fine-tuning on Qwen2.5-7B-Instruct with the LLaMA-Factory framework. 3. Multi-agent design: Implemented an end-to-end multi-agent framework for automated CFD simulations based on MetaGPT.

🛠️ PROFESSIONAL SKILLS

- **Programming Languages:** Python, C++, Java
- **English:** CET-6 (518)