

Dairu Liu

ironmt00@gmail.com · dairuli.github.io · Beijing, China · +86-175-0102-4327

EDUCATION

Nankai University
B.Eng. in Software Engineering

Tianjin, China
Sep. 2023 – Jun. 2027 (expected)

The Affiliated High School of Peking University

Beijing, China

RESEARCH INTERESTS

My primary research goal is to leverage large-scale data to build **generalizable** embodied agents capable of performing diverse real world tasks. I am particularly interested in **Robot Learning** (scaling up robot data, efficient robot learning algorithms, humanoid whole-body control, loco-manipulation) and **Multimodal Reasoning** (agentic perception, cross-modal understanding, long-horizon decision making).

PUBLICATIONS & PREPRINTS

[1] **HumanTracker: Towards Comprehensive and Human-Aligned Motion Tracking Benchmark**

Dairu Liu*, Zekun Qi*, Jiayu Zeng*, Yu Guan, Chenghuai Lin, Xuchuan Chen, Xinqiang Yu, Wen Yao Zhang, He Wang[†], Li Yi[†]

ECCV 2026

- Built a standardized humanoid motion tracking benchmark with **150 hours** of newly captured optical mocap from **24 professional performers**, comprising **24K clips** organized into **4 motion families** for fine-grained diagnosis.
- Proposed **HumanScore**, a preference-aligned reward model trained on **3K pairwise human comparisons**; achieves **0.8834** alignment rate with human judgments, outperforming MPJPE (**0.8164**), and validated on **Unitree G1** real-robot deployment.

[2] **Humanoid-GPT: Scaling Data and Structure for Zero-Shot Motion Tracking**

Zekun Qi*, Xuchuan Chen*, Dairu Liu*, Chenghuai Lin*, Yunrui Lian, Zhikai Zhang, Wen Yao Zhang, Yu Guan, Jilong Wang, Xinqiang Yu, He Wang[†], Li Yi[†]

CVPR 2026 [Paper] [Project] [Code]

- Introduced **Humanoid-GPT**, a GPT-style causal Transformer that scales humanoid motion tracking through large-scale data and structured policy learning, enabling robust zero-shot motion tracking on **Unitree G1**.
- Distilled hundreds of RL motion experts into one generalist tracker using DAgger and proposed **Harmonic Motion Embedding (HME)** for diversity-aware, distribution-balanced sampling over a **~2B-frame** retargeted motion corpus.

[3] **LIMMT: Less is More for Motion Tracking**

Yu Guan*, Zekun Qi*, Chenghuai Lin, Xuchuan Chen, Dairu Liu, Wen Yao Zhang, Jilong Wang, Xinqiang Yu, He Wang[†], Li Yi[†]

ICML 2026 [Paper]

- Conducted the first data-centric study for humanoid motion tracking, showing that carefully selected compact motion subsets can outperform training on the full motion corpus.
- Proposed **General Quality Selection (GQS)**, which scores motions by physics feasibility, diversity, and complexity, then selects compact subsets via complexity-weighted farthest-point sampling; training on only **~3%** of AMASS outperforms the full corpus across multiple trackers and transfers to real-world **Unitree G1** deployment.

[4] **Thinking with Visual Abstract: Enhancing Multimodal Reasoning via Visual Abstraction**

Dairu Liu*, Ziyue Wang*, Minyuan Ruan*, Fuwen Luo, Chi Chen, Peng Li[†], Yang Liu[†]

NeurIPS 2025 Workshop [Paper] [Project]

- Proposed **Visual Abstract Thinking (VAT)**, a reasoning paradigm that replaces verbose verbal CoT with visual abstractions, focusing models on semantic, structural, and geometric cues.
- VAT achieves **+2.21%** average gain over the GPT-5 baseline, exceeding CoT (**+0.96%**); uses only **0.68×** CoT runtime and **0.1×** the tokens of Visual Sketchpad on GPT-4o. RL-trained VAT further gains **+4.52%** on Qwen-2.5-VL-3B.

[5] **Evaluating Time Awareness and Cross-modal Active Perception of Large Models via 4D Escape Room Task**
Yurui Dong*, Ziyue Wang*, **Dairu Liu***, Shuyun Lu*, Xuechen Liu*, Fuwen Luo, Peng Li†, Yang Liu†
In Submission [\[Paper\]](#) [\[Project\]](#)

- Developed **EscapeCraft-4D**, a customizable 4D escape-room benchmark (**66 scenes, 1,032 objects**) incorporating trigger-based audio, misleading auditory distractors, and temporally transient clues for evaluating time-aware multimodal reasoning.
- Introduced 4D specific process level metrics beyond final success rate: **AMR** (Audio Misleading Rate, measuring how often agents are misled by auditory distractors) and **TCSS** (Transient Clue Seeing Speed, measuring how early agents discover irreversibly-disappearing clues), revealing significant modality bias and failures in time-conditioned decision making across Omni models.

RESEARCH EXPERIENCE

Research Intern, GALBOT Advisor: Prof. Li Yi, Prof. He Wang	Beijing, China Dec. 2025 – Present
• Built a large-scale humanoid motion tracking benchmark and proposed a preference-aligned reward model for whole-body control evaluation, addressing the misalignment between kinematic metrics and human perception. [1] • Contributed to Humanoid-GPT, a large-scale generalist humanoid motion tracker; responsible for extensive real-robot experiments on Unitree G1 and large-scale motion data processing for training and evaluation. [2]	
Research Intern, THUNLP AIR, Tsinghua University Advisor: Prof. Peng Li, Prof. Yang Liu	Beijing, China Apr. 2024 – Jan. 2026
• Proposed Visual Abstract Thinking, a novel multimodal reasoning paradigm that enhances MLLMs via visual abstraction rather than verbose verbal rationales. [4] • Developed a 4D escape-room benchmark for evaluating time awareness and cross-modal active perception in Omni models. [5]	

HONOURS & AWARDS

National Scholarship (Top 0.2%; Undergraduate, 10000CNY)	Sep. 2024
Second Prize , NOIP 2018 (National Olympiad in Informatics in Provinces)	Dec. 2018
Second Prize , CSP-S 2019 (CCF Certified Software Professional)	Dec. 2019

LEADERSHIP & SERVICES

Vice President, Quantitative Finance & Investment Society, Nankai University	Sep. 2025 – Present
Member, University Ski Team, Nankai University	2024 – 2025
Academic Department Vice-Head, Psychology Association, Nankai University	Sep. 2024 – Sep. 2025

SKILLS

Robot Platforms: Unitree G1
Programming: C++, Python, MATLAB, L^AT_EX
Languages: Mandarin Chinese (native); English (IELTS 7.0 – L 7.5, R 8.0, S 6.0, W 6.0; CET-6 608)
Interests: Skiing