Bohan Zhou

+86 15797895657 | zhoubh@link.cuhk.edu.hk | Homepage

EDUCATION

The Chinese University of Hong Kong Hong Kong, China
Ph.D., Department of Computer Science and Engineering *Aug. 2026 – Present*

- Research: **embodied intelligence and foundation models**. One published paper in a CCF-A venue.
- **3rd Prize, Gripper Track**, the third X Square–Tsinghua Embodied AI Challenge.

Peking University Beijing, China
*Master's degree in Computer Applied Technology; GPA: **3.88/4.0*** *Sep. 2023 – Jul. 2026*

- National Engineering Research Center of Visual Technology. Research: **multimodal learning, reinforcement learning and agents**.
- **12 published papers**, including **4 first-author papers** and **8 in CCF-A venues**.

Nankai University Tianjin, China
Bachelor's degree in Intelligent Science and Technology, College of Artificial Intelligence *Sep. 2019 – Jul. 2023*

- GPA: **3.95/4.0**; rank: **2/98**. Recipient of seven undergraduate scholarships.

INTERNSHIP EXPERIENCE

Beijing Academy of Artificial Intelligence (BAAI)
Research Internship — Generalist and Embodied Agents *Sep. 2022 – Apr. 2025*

- Advanced **general embodied intelligence from virtual to physical environments** through video-driven policy learning, visual navigation and multimodal dexterous manipulation: STG, NOLO, MEgoHand and BiDexHD.
- Produced **4 first-author papers**: NeurIPS 2023/2025, IROS 2025 Oral and AAAI 2026 Oral. STG, NOLO and MEgoHand were featured by Prof. Zongqing Lu at the **BAAI Conference in three consecutive years, 2023–2025**.

JD Explore Academy
TGT Talent Program *May 2025 – Jul. 2025*

- **EgoLive**: Helped build the most comprehensively annotated open manipulation-focused egocentric dataset: **1,680 hours, 65K episodes, 346 tasks**; $\sim$**2**$\times$ **EgoDex** and $\sim$**1.6**$\times$ **Xperience-10M** in scale, covering unconstrained homes, retail and pharmacies.
- **Annotation**: Integrated JoyEgoCam stereo RGB/**200 Hz IMU**, calibration/rectification/alignment, HaMeR hands, ORB-SLAM3 6-DoF poses, FoundationStereo depth and Qwen3-VL-32B semantics.
- **VLA-H2**: Transferred MEgoHand human hand–object pretraining to dexterous-hand policy post-training across scenes, adding **object modeling, cross-embodiment retargeting and action-trajectory optimization**.

Momenta
M-star Talent Program *Aug. 2025 – Nov. 2025*

- **Data**: Built a scenario-mining and cleaning loop, raising intersection-action recognition from approximately **70% to 97%** and reducing annotation costs by **51%–62%**.
- **Models**: Coordinated **R6 post-training at the 100-GPU scale**; led **Flow Matching trajectory generation and FPO safety enhancement**, outperforming the previous autoregressive baseline in simulation and road tests.
- **Deployment**: Delivered operator optimization, export, closed-loop simulation and **30 km of IM LS6 testing on Beijing's Fourth Ring Road** at $>$**10 Hz**. The FM approach became the mainline and entered current products.

BeingBeyond

Embodied Foundation Models and Dexterous Systems *Dec. 2025 – Feb. 2026*

- **Being-H0**: Contributed the core human-video data pipeline, combining **MEgoHand, TASTE-Rob and EgoDex** to learn general manipulation patterns; connected human-video pretraining to robot-policy transfer through physical instruction tuning.
- **RealDexUMI**: Helped develop a low-cost wearable dexterous data-collection system with **isomorphic control and integrated vision–tactile end effectors**, connecting demonstration capture to deployable fine manipulation across embodiments.

KnowinAI

Robot Deployment and Embodied Skill Coordination *Mar. 2026 – Aug. 2026*

- **Robot deployment**: Delivered Astra simulation/real-robot evaluations and dual-arm dexterous systems, integrating perception, control and hardware. Used OpenCV tracking, ramped acceleration and handovers for responsive interaction and automatic stowing.
- **RoboAware, Project Lead**: Proposed **System 2 Coding Agent + System 1 Mod/E2E** coordination. Learned state-dependent skill assignment from same-state counterfactual branches using **MCTS and Q-learning**, with low-level executors frozen. *ICLR 2027 submission: mean review score 2.67, top 15%.*

PUBLICATIONS

[1] **Bohan Zhou**, Ke Li, Jiechuan Jiang, et al. Learning from visual observation via offline pretrained state-to-go transformer. *NeurIPS 2023*.

[2] **Bohan Zhou**, Yi Zhan, Zhongbin Zhang, et al. MEgoHand: Multimodal Egocentric Hand-Object Interaction Motion Generation. *NeurIPS 2025.*

[3] **Bohan Zhou**, Zhongbin Zhang, Jiangxing Wang, et al. NOLO: Navigate Only Look Once. *IROS 2025 Oral.*

[4] **Bohan Zhou**, Haoqi Yuan, Yuhui Fu, et al. Learning diverse bimanual dexterous manipulation skills from human demonstrations. *AAAI 2026 Oral.*

[5] Haoqi Yuan, **Bohan Zhou**, Yuhui Fu, et al. Cross-embodiment dexterous grasping with reinforcement learning. *ICLR 2025.*

[6] Sipeng Zheng, **Bohan Zhou**, Yicheng Feng, et al. UniCode: Learning a unified codebook for multimodal large language models. *ECCV 2024*.

[7] Hao Luo, **Bohan Zhou**, Zongqing Lu. Pre-trained Visual Dynamics Representations for Efficient Policy Learning. *ECCV 2024*.

[8] Weihao Tan, Wentao Zhang, Xinrun Xu, Haochong Xia, Ziluo Ding, Boyu Li, **Bohan Zhou**, et al. Cradle: Empowering foundation agents towards general computer control. *ICML 2025.*

[9] Zhengxi Hu, Yuxue Yang, Xiaolin Zhai, Dingye Yang, **Bohan Zhou**, et al. GFIE: A Dataset and Baseline for Gaze-Following from 2D to 3D in Indoor Environments. *CVPR 2023*.

[10] Zhengxi Hu, ..., **Bohan Zhou**, et al. Gaze target estimation inspired by interactive attention. *TCSVT 2022.*

[11] Qize Yu, ..., **Bohan Zhou**, et al. GroundingPI: A Grounding Foundation Model towards Physical Intelligence with Visual Primitives. *(Co-author; Under Review; reached Hugging Face Daily Papers #1).*

[12] Qize Yu, ..., **Bohan Zhou**, ..., et al. AffordanceVLA: A Vision-Language-Action Model Empowering Action Generation through Affordance-Aware Understanding. *ECCV 2026 Workshop*.

[13] Haoqi Yuan, Yu Bai, Yuhui Fu, **Bohan Zhou**, et al. Being-0: A Humanoid Robotic Agent with Vision-Language Models and Modular Skills. *(Under Review).*

[14] **Bohan Zhou**, et al. FETCH: Frame-grounded Evidence Topology for Embodied Spatial Planning with the Counterfactual Hypothesis. *(Project Lead, Under Review).*

[15] **Bohan Zhou**, et al. RoboAware: Learning to Coordinate Embodied Skills from Counterfactual Outcomes. *(Project Lead, ICLR 2027 submission).*

SELECTED RESEARCH CONTRIBUTIONS

BiDexHD: Scalable Bimanual Dexterous Learning AAAI 2026 Oral
First Author *Jun. 2024 – Oct. 2024*

- Overcame task-specific reward engineering by **automatically constructing robot tasks from human demonstrations** and learning vision-based bimanual skills with a shared reward design.
- Scaled to **141 tool-use tasks**; achieved **51.07%** and **65.99%** success on unseen TACO and ARCTIC tasks, respectively.

MEgoHand: Human Hand–Object Interaction VLA NeurIPS 2025
First Author *Mar. 2025 – May 2025*

- Introduced a **human HOI vision–language–action framework**, pioneering VLM-based motion priors for egocentric interaction. Generated **future hand motions** beyond observed-pose reconstruction while reducing dependence on predefined object models.
- Contributed **3M+ HOI training frames** for generalization to unseen objects and across datasets; invited to present at **3D Vision Workshop**.

STG: Offline Video Pretraining for Visual Control NeurIPS 2023
First Author *Sep. 2022 – May 2023*

- Introduced **State-to-Go latent-transition pretraining** to address inefficient online visual imitation and restrictive predefined goals, learning scalable behavioral priors from offline video alone.
- Converted video priors into intrinsic rewards for **online reinforcement learning without environment rewards**, improving learning on Atari and Minecraft.

NOLO: Navigate Only Look Once IROS 2025 Oral
First Author *Feb. 2024 – May 2024*

- Enabled **one-shot, in-context adaptation to new scenes** for image-goal navigation using one **30-second scene video** and online egocentric RGB, without maps, depth, pose inputs or retraining.
- Validated maze navigation on a **Unitree Go2** and received an invitation to present at **3D Vision Workshop**.

Cradle: General Computer Control ICML 2025
Core Contributor *Oct. 2023 – Feb. 2024*

- Pioneered **General Computer Control (GCC)**, an early computer-use-agent paradigm unifying screenshot input and keyboard/mouse output across software. Completed approximately **40-minute storyline missions in Red Dead Redemption 2**.
- Implemented **low-level keyboard/mouse interaction and an evolving skill library**, organizing reusable skills by name, parameters and functionality for generation, refinement and execution.
- Reached **2.6K+ GitHub stars**; the follow-on VARP framework reused core modules for *Black Myth: Wukong*.

GroundingPI: A Grounding Foundation Model Under Review
Co-author — Architecture, Training and Evaluation *2026*

- Strengthened fine-grained localization through **visual primitives**, improving downstream manipulation generalization and data efficiency. The **4B** model led the aggregate comparison across **34 grounding benchmarks**, averaging **73.68%** against 44 baselines.
- Trained on a **512-GPU B300 cluster**, using **128 GPUs per stage** and **721.58B tokens** of cumulative training exposure; reached **Hugging Face Daily Papers #1**.

FETCH: Frame-grounded Evidence Topology Under Review
First Author — Embodied Spatial Planning

- Turned isolated spatial relations into reliable action orderings through **physical-reference-frame evidence topology and counterfactual necessity verification**, producing structured plans directly usable by controllers.
- Introduced **Typed Multi-Token Prediction** for **3.78× faster decoding**; achieved **90% real-robot success** on a conditional table-clearance task.

RoboAware: Outcome-aware Skill Coordination Under Review
First Author and Project Lead *ICLR 2027 submission*

- Led a team to learn **state-dependent coordination between modular skills and end-to-end policies** from **same-state counterfactual outcomes**, keeping low-level executors frozen.
- Achieved **77.0% average success across reported evaluation splits covering 100 tasks** on three benchmarks; demonstrated **zero-shot deployment on long-horizon real-robot tasks**.

ADDITIONAL RESEARCH EXPERIENCE

Being-0: A Humanoid Robotic Agent with Vision-Language Models and Modular Skills
Beijing Academy of Artificial Intelligence, Embodied AI *Oct. 2024 – Feb. 2025*

- **Hierarchical humanoid control**: Combined a black-box LMM for task decomposition, a finetuned LMM for navigation and low-level dexterous skills.
- **VLM/rule-based coordination**: Connected language plans to motor skills for locomotion, navigation and dexterous manipulation.
- **Deployment**: Completed **tens-of-meters navigation and coffee delivery on Unitree H1**.

PVDR: Pre-trained Visual Dynamics Representations for Efficient Policy Learning
Beijing Academy of Artificial Intelligence, Embodied AI *2024*

- **PVDR [ECCV 2024]**: Proposed pretraining visual transfer representations from large-scale unlabeled Internet videos to overcome domain gaps and adapt to downstream reinforcement learning tasks.

UniCode: Learning a Unified Codebook for Multimodal Large Language Models
Beijing Academy of Artificial Intelligence, Generalist Agents *2024*

- **UniCode [ECCV 2024]**: Proposed an efficient multimodal training paradigm that learns a unified codebook, relaxing the constraints of pure-text codebooks on image generation.

Human Intent Analysis in Indoor Environments for Service Robots Advisor: Prof. Jingtai Liu
Tianjin Key Laboratory of Intelligent Robotics *Mar. 2021 – Aug. 2022*

- **Human Intent Analysis Pipeline**: Developed semi-automatic data construction and gaze direction/target estimation for human-intent understanding and prediction; received **3rd Prize**.
- **GFIE [CVPR 2023]**: Combined Kinect and laser ranging for gaze collection and unbiased 2D/3D target annotation through laser-spot localization.
- **VSG-IA [TCSVT 2022]**: Developed graph attention for gaze behavior detection and human–scene interaction analysis.

AWARDS & HONORS

- **Beijing Outstanding Graduate**, Peking University 2026
- **National Scholarship; Outstanding Student Award**, Peking University Sep. 2025
- **Tianjin Municipal Outstanding Graduate**, Nankai University Jun. 2023
- **Tianjin Municipal People's Government Scholarship (top 2%)** Nov. 2021
- **2nd Prize**, National College Student Mathematical Modeling Competition Oct. 2021
- **Honorable Mention**, American College Mathematical Contest in Modeling May 2021
- **Gongneng Scholarship (top 5%)**, Nankai University Dec. 2020, Dec. 2022

SKILLS

- **Technical**: PyTorch; distributed training on 100 GPUs; hands-on coding and AI-assisted development.
- **Research and delivery**: Project planning, team coordination, technical communication and on-time delivery.
- **English**: CET-6 score **622**; speaking grade **A**.