

Jihoon Hong

+1-470-667-7557 | jhong392@gatech.edu

[Website](#) | [Google Scholar](#) | [LinkedIn](#)

Atlanta, GA

RESEARCH PROFILE

Robotics and embodied-AI researcher developing robust world-action models and provably safe perception-control systems for robot manipulation and navigation using activation steering, optimal control, and compact 3D scene representations.

EDUCATION

- **Georgia Institute of Technology** Aug. 2024 - Expected Aug. 2029
Ph.D., Machine Learning Atlanta, GA
 - Advisors: Dr. Sara Fridovich-Keil and Dr. Glen Chou
- **Seoul National University** Mar. 2017 - Feb. 2024
B.A., Economics & B.S., Computer Science and Engineering Seoul, Korea
 - *Summa Cum Laude* (GPA: 3.93/4.0)
 - Minor in Mathematics
 - Mandatory military service, Republic of Korea Army (Aug. 2019 - Feb. 2021)

TECHNICAL EXPERTISE

- **Embodied AI:** World action models, robot learning, VLA model architectures, robustness under distribution shift, LIBERO-10
- **Control and Robotics:** LQR, CBFs, motion planning, safe navigation, Crazyflie hardware deployment
- **Perception and Model Analysis:** 3D Gaussian splatting, DiTs, activation steering, mechanistic interpretability
- **Languages:** Python, C++ | **ML/GPU:** PyTorch, CUDA, Hugging Face Diffusers | **Robotics:** MuJoCo

SELECTED RESEARCH PROJECTS

- **Robust World Action Models (WA-LQR) [Paper]** Under review at CoRL 2026
Trustworthy Robotics Lab — Advisor: Glen Chou
 - Led development of a training-free activation-steering framework that identifies low-dimensional robustness features in world action models and synthesizes closed-loop interventions using local linear dynamics and LQR.
 - Evaluated NVIDIA Cosmos-Policy 2B, DiT4DiT, and LingBot-VA on LIBERO-10; improved task success rate by up to 41 percentage points under camera, gripper-position, and sensor perturbations without fine-tuning.
- **Controllable and Safe Video Generation (LA-LQR) [Paper]** Under review at NeurIPS 2026
Trustworthy Robotics Lab — Advisor: Glen Chou
 - Developed a closed-loop controller that projects DiT activations into contrastive subspaces, estimates dynamics with Jacobian-vector products, and computes LQR interventions.
 - Reduced average safety-violation rates from 36.9% to 14.6% on Wan2.1-T2V-14B/T2VSafetyBench and from 31.5% to 12.2% on HunyuanVideo-1.5/SafeSora while preserving subject consistency.
- **Safe Navigation with 3D Gaussian Splats (PolyMerge) [Paper]** IEEE RA-L 2026 (IROS 2026)
VOILA and Trustworthy Robotics Labs — Advisors: Sara Fridovich-Keil and Glen Chou
 - Designed a provably conservative convex-polytope compression of 3D Gaussian-splat scenes for integration with control-barrier-function safety filters.
 - Achieved up to 10× lower memory use and approximately 100× faster safety-filter computation; deployed onboard a Crazyflie 2.1+ at 125–500 Hz.
- **Diffusion-Based OOD Detection (KLIP & LACToS)** KLIP: CVPR 2026 | LACToS: Under review at NeurIPS 2026
VOILA Lab — Advisor: Sara Fridovich-Keil
 - Developed KLIP [Paper], a prior-posterior KL-divergence metric that detects and localizes OOD features under inverse problem settings.
 - Developed LACToS, a typicality-based OOD detection framework using superpixels; evaluated across industrial defects, CT tumors, facial anomalies, and X-ray security screening.

PUBLICATIONS & MANUSCRIPTS

* denotes equal contribution.

Peer-Reviewed Publications

- J. Hong, C. Chiu, S. Fridovich-Keil, G. Chou. **PolyMerge: Compressing 3D Gaussian Splats with Polytope Coverings for Provably Safe Resource-Constrained Navigation**. *IEEE RA-L*, 2026; accepted for presentation at IROS 2026.
- A. Kheirandish*, J. Hong*, S. Fridovich-Keil. **KLIP: localized distribution shift detection via KL-divergence with diffusion priors in Inverse Problems**. *CVPR* 2026.
- Z. Ye, K. Xia, Y. Fu, X. Dong, J. Hong, X. Yuan, S. Diao, J. Kautz, P. Molchanov, Y. Lin. **LongMamba: Enhancing Mamba's Long Context Capabilities via Training-Free Receptive Field Enlargement**. *ICLR* 2025.
- Z. Ye, Z. Wang, K. Xia, J. Hong, L. Li, L. Whalen, C. Wan, Y. Fu, Y. Lin, S. Kundu. **LAMB: A training-free method to enhance the long-context understanding of SSMs via attention-guided token filtering**. *ACL* 2025.
- Z. Ye, C. Wan, C. Li, J. Hong, S. Li, L. Li, Y. Zhang, Y. Lin. **3D Gaussian Rendering Can Be Sparser: Efficient Rendering via Learned Fragment Pruning**. *NeurIPS* 2024.
- S. Lee, H. Lee, J. Hong, S. Cho, J. Lee. **VGA: Hardware Accelerator for Scalable Long Sequence Model Inference**. *MICRO* 2024.
- H. Lee, J. Hong, S. Kim, S. Lee, J. Lee. **A Memory-Efficient Edge Inference Accelerator with XOR-based Model Compression**. *DAC* 2023.

Manuscripts Under Review

- J. Hong*, J. Skifstad*, Q. Dai, A. Chan, G. Chou. **Steering Robustness into World Action Models via Mechanistic Interpretability and Optimal Control**. *Under review at CoRL* 2026.
- J. Hong, A. Chan, Q. Dai, J. Skifstad, G. Chou. **Activation Steering of Video Generation Models via Reduced-Order Linear Optimal Control**. *Under review at NeurIPS* 2026.
- J. Hong*, A. Kheirandish*, S. Fridovich-Keil. **LACToS: Local Anomaly detection via Conditional Typicality of Superpixels**. *Under review at NeurIPS* 2026.

HONORS AND AWARDS

- Kwanjeong Educational Foundation (KEF) Scholarship

Sep. 2024 – Aug. 2029