YUZHEN CHEN

+1 734-548-0546 ◇ yuzhenchen@g.harvard.edu ◇ yuzhench.github.io ◇ @yuzhench

EDUCATION

Harvard University Cambridge, MA
M.S. in Computational Science and Engineering Expected May 2027
Relevant Coursework: Video Generation & Multi-modality World Model & VLA

University of Michigan Ann Arbor, MI
B.S.E. in Computer Engineering; **GPA:** 3.9/4.0 May 2025
Relevant Coursework: Data Structures & Algorithms, Machine Learning, Computer Vision, Computer Architecture, Robotic Kinematics and Dynamics, Robot Perception, SLAM, Linear Algebra, Multivariable Calculus, Differential Equations, Discrete Math, Probability and Statistics, Digital Circuit Design, Logic Design, Embedded System, GPU Parallel Computation, Robot 3D Reconstruction, Human Computer Interactive

PUBLICATIONS

* denotes equal contribution.

1. **Chen, Y.**, Zhou, K. (2026). WCM: World-Cognition Model for Generalizable Human–Robot Interaction. *RSS 2026 Workshop on Human-centric Mobile Manipulation.*
2. Liu, Y.*, **Chen, Y.***, Wang, B.*, Li, C., Lu, J., Li, H., Zhou, K., Wang, M., Zhan, F. (2026). Act to See: Structured Articulated Representations through Robot Interaction. *RSS 2026 FM4RoboPlan Workshop.*
3. Zhou, K.*, **Chen, Y.***, Zhan, F., Hua, H., Chen, G., Chang, X., Qu, A., Du, Y., Liu, Z., Liang, P.P., Wang, M. (2026). GEM-4D: Geometry-Enhanced Video World Models for Robot Manipulation. *Under review, ECCV 2026.* **(Best Paper Runner-Up & Spotlight, CVPR 2026 Workshops)**
4. Zhou, K.*, **Chen, Y.***, Yildiz, E.*, Shi, M., Dai, D., Chen, G., Zheng, J., Wang, H., Zhan, F., Saini, C., Shen, L.Q., Guo, Y., Liang, P.P., Wang, M. (2026). Learning Ophthalmologist Clinical Reasoning for Glaucoma Diagnosis from Fundus Images. *medRxiv preprint* (under review, *npj Digital Medicine*).
5. Chen, Y., Wu, X., Zong, Y., **Chen, Y.**, Li, A., Wu, J., Zhang, B., Vasudevan, R. (2025). DEFT: Differentiable Branched Discrete Elastic Rods for Modeling Furcated DLOs in Real-Time. *arXiv, 2025.*
6. **Chen, Y.**, Son, H., Kusari, A. (2025). MatPredict: A Dataset and Benchmark for Learning Material Properties of Diverse Indoor Objects. *arXiv preprint* (under review).
7. Chen, Y., Zhang, Y., Brei, Z., Zhang, T., **Chen, Y.**, Wu, J., Vasudevan, R. (2024). Differentiable Discrete Elastic Rods for Real-Time Modeling of Deformable Linear Objects. *CoRL 2024*, Munich.
8. Ewen, P., Gunjal, G., Chen, H., Li, A., **Chen, Y.**, Vasudevan, R. (2024). Multi-Modal Semantic Perception Using Bayesian Inference. *IROS 2023*, Detroit. **(Best Poster Award)**
9. Ewen, P., Chen, H., **Chen, Y.**, Li, A., Bagali, A., Gunjal, G., Vasudevan, R. (2024). You've Got to Feel It To Believe It: Multi-Modal Bayesian Inference for Semantic and Property Prediction. *RSS 2024.*
10. Sheppard, A., Ewen, P., Wilson, J., Sethuraman, A.V., Adewole, B., Li, A., **Chen, Y.**, Vasudevan, R., Skinner, K.A. (2026). SLIM-VDB: A Real-Time 3D Probabilistic Semantic Mapping Framework. *IEEE RA-L.*
11. Oral Presentation: **Chen, Y.**, Cai, B. (2023) Development of high-performance but low-cost sensing node for intelligent monitoring system. *U-M UROP Symposium*, Ann Arbor, MI. **(Best Poster Award)**

RESEARCH EXPERIENCE

GEM-4D: Geometry-Enhanced Video World Models for Robot Manipulation [3] Cambridge, MA
Research Assistant, Advisors: Dr. Paul Pu Liang (MIT Media Lab) & Dr. Mengyu Wang (Harvard AI and Robotics Lab) Sept. 2025 - Present

- Developed a geometry-grounded video world model that distills dense 4D correspondence supervision into a video diffusion transformer during training, improving physical consistency with no added inference cost.
- Proposed an adaptive inverse dynamics system that converts generated videos into executable 6-DoF robot trajectories using vision foundation models and confidence-gated pose recovery.

- Achieved state-of-the-art 4D scene generation—improving RGB video quality (SSIM 76.2→82.1, FVD 33.2→31.8) and point-cloud reconstruction (Chamfer distance 0.244→0.200) over the best prior baselines—and raised real-world robot manipulation success from 61% to 81%.

WCM: World-Cognition Model for Generalizable Human–Robot Interaction [1] Cambridge, MA
First Author & Project Lead, Harvard University Feb. - July 2026

- Designed WCM, a human-centered embodied agent built on the SLAK architecture (Sensing, Logic, Action, Knowledge) with an asynchronous runtime that lets reasoning, dialogue, and execution proceed concurrently, so the robot keeps acting instead of halting to think between steps.
- Introduced a human-in-the-loop teaching mode and a chain-of-thought distillation pipeline that refines both teaching episodes and autonomous rollouts into step-level CoT supervision, enabling the agent to acquire long-horizon tasks from user demonstration and improve continually.
- Achieved a 73.8% average success rate across a nine-task real-world human–robot interaction suite—including tasks held out from CoT fine-tuning and a long-horizon task learned purely through teaching—while cutting task time from 103.5 s to 38.5 s and GPU busy time from 103.5 s to 22.0 s against a VLA baseline.

Learning Ophthalmologist Clinical Reasoning for Glaucoma Diagnosis from Fundus Images [4] Boston, MA
Research Assistant, Advisor: Dr. Mengyu Wang (Harvard AI and Robotics Lab) Jan. 2026 - Present

- Built the first expert-annotated glaucoma reasoning dataset—1,077 fundus photographs from the LAG and PAPILA databases paired with ophthalmologist-authored six-step diagnostic reports covering image quality, optic disc, cup-to-disc ratio, neuroretinal rim, and ISNT assessment.
- Developed a reasoning-driven vision-language framework that generates a structured clinical reasoning trace *before* the diagnosis on a shared RETFound backbone, making the prediction interpretable and clinically traceable rather than a direct black-box classification.
- Achieved 94.7% balanced accuracy and 94.8% precision, outperforming retinal foundation models (RetiZero 93.6%, RETFound 93.1%) and fine-tuned medical VLMs; generated reports were clinically validated at 0.070 cup-to-disc-ratio MAE, 1.73 ISNT Kendall distance, and 0.874 BERTScore-F1 against expert reports.

Act to See: Structured Articulated Representations through Robot Interaction [2] Boston, MA
Research Assistant, Advisor: Dr. Mengyu Wang (Harvard AI and Robotics Lab) Jan. 2026 - Present

- Treated robot interaction as an online reconstruction signal rather than a one-off offline measurement: each streaming RGB-D frame is fused into an explicit URDF *while* the part is still being moved, recovering joint type, axis, and position primarily from the end-effector trajectory with sparse visual tracks as fallback.
- Designed an axis-triggered rollback mechanism that detects when incoming frames stop agreeing with the current articulation hypothesis, rewinds to cached frames, and re-arbitrates the prismatic/revolute fits—correcting early mis-typing within the same interaction instead of committing to it.
- Reached a 100% valid reconstruction rate on both joint types in SAPIEN (vs. 76.9% for prior methods) with 2.60° prismatic axis error, converging 4.2 s after manipulation ends versus 182–874 s for offline baselines; validated closed-loop on a real UFactory xArm 6 with a single third-person RGB-D camera, and showed the resulting URDF is directly reusable as memory for a VLM-based planner.

Multi-Modal Semantic Perception in Robotic Navigation [8], [9] Ann Arbor, MI
Research Assistant, Advisor: Dr. Ram Vasudevan Apr. 2023 - May 2025

- Modified Kinect Fusion to integrate material-based color mapping into 3D reconstruction, enabling real-time visualization of surface friction coefficients.
- Optimized material classification model using MiT-B1 architecture, achieving 4x faster inference (9 FPS) while maintaining accuracy across 6 material classes.
- Developed a voting pipeline combining SAM and SegFormer models, significantly reducing high-noise issues at object junctions in real-time material predictions.

Learning Physical Object Properties from Images Using Diffusion Model [6] Ann Arbor, MI
Research Assistant, Advisor: Dr. Arpan Kusari Sept. 2024 - May 2025

- Developed an automated pipeline using Meta MatSynth and Replica datasets to generate large-scale training data with varied material properties, lighting conditions, and camera angles.
- Trained an encoder to extract material properties (base color, metallic, roughness, reflectance, anisotropy, transmission) from images, enabling accurate Physically-Based Rendering function generation.
- Implemented Monte-Carlo differentiable rendering as decoder, achieving robust image reconstruction.

Conformalized Reachable Sets for Obstacle Avoidance With Spheres (CROWS) Ann Arbor, MI
Research Assistant, Advisor: Dr. Ram Vasudevan Apr. 2024 - Sept. 2024

- Developed ML models for CROWS, a safe motion planning system using conformalized neural networks for sphere parameter prediction, reducing planning time from 0.51s to 0.21s compared to ARMTD.
- Generated extensive training datasets through simulation-based sampling of motion trajectories, enabling robust training of three distinct neural networks for center, radii, and gradient predictions.
- Validated system effectiveness through successful collision-free motion demonstrations in both simulation and hardware experiments, ensuring safe robot operation in unstructured environments.

Real-Time Modeling for Elastic Rods [5], [7] Ann Arbor, MI
Research Assistant, Advisor: Dr. Ram Vasudevan Nov. 2023 - Sept. 2024

- Used DEFORM, a physics-based modeling system for real-time simulation of elastic rods, achieving accurate prediction of rope configurations in 3D space based on physical parameters.
- Implemented a motion capture pipeline for collecting ground truth data, facilitating model training, validation, and performance optimization through real-world rope behavior analysis.
- Applied the simulation framework to industrial robotics, enabling automated wire manipulation tasks including insertion, removal, and assembly operations.

SLIM-VDB: Real-Time 3D Probabilistic Semantic Mapping [10] Ann Arbor, MI
Research Assistant, Advisor: Dr. Katie Skinner May 2024 - Sept. 2024

- Contributed to a real-time 3D probabilistic semantic mapping system that leverages OpenVDB for memory-efficient volumetric scene representation.
- Integrated closed-set and open-set semantic predictions with RGB-D / point-cloud observations through probabilistic fusion to generate interpretable 3D semantic maps.
- Built visualization and evaluation tools for benchmark mapping experiments, supporting analysis of semantic quality, integration time, and memory efficiency.

INTERNSHIPS

Bardor | Software & System Engineer May - Sept. 2025

- Bridging software/hardware requirements between Detroit clients and Chinese manufacturers, securing projects for instrument clusters and infotainment displays across Chrysler, Jeep, and Polaris platforms.

UMich Shapiro Design Lab | Educational Technology Developer May 2024 - May 2025

- Designed and built a multi-screen dice with six independent displays that respond to motion, enabling dynamic group learning in language education and serving as a platform for game design exploration - deployed at university language centers and gaming design courses, presented at Serious Play Conference 2025.

PROJECTS

Embodied Intelligence Robot – ROVER | Independent Project May. 2025 - Present

- Engineered a high-performance 5-DOF robotic arm achieving 1-meter reach with 1kg payload capacity, delivering 3x better cost-effectiveness (<$2000) than commercial alternatives
- Achieved 50ms control loop latency through optimized ROS2 architecture and custom motion planning algorithms, enabling real-time multi-modal interaction with RGB-D SLAM, LiDAR, and natural language processing
- Served as a testing platform for research teams, accelerating VLA/VLM model deployment cycles through providing affordable, reproducible physical validation beyond simulation

HONORS & AWARDS

OMPP Summer Research Fellowship | Harvard University 2026
Best Paper Award, Runner-Up | CVPR 2026 WMAS Workshop 2026
Spotlight | CVPR 2026 3D-LLM/VLA Workshop 2026
Best Poster Award | IROS 2023 2023
James B. Angell Scholar | University of Michigan 2023-2024
Dean's List | University of Michigan 2022-2025

LEADERSHIP POSITIONS

Junior Organizer | PhysWorldAI: 1st Workshop on Physical World AI, NeurIPS 2026 2026
Robotics Community Event Organizer | MIT CHIEF 2026 - Present
Teaching Assistant | EECS442 Computer Vision; Physics 140X Sept. 2022 - Apr. 2025
Departmental Ambassador | ENGR 110; UMich Robotics Department Sept. 2023 - Dec. 2024
Student Tutor | UMich Engineering Center for Academic Success Jan. - Apr. 2023

SKILLS

Programming Languages: Python, C++, C, CUDA, MATLAB, JavaScript, Shell Scripting, LaTeX
ML/DL Frameworks: PyTorch, TensorFlow, JAX, TensorRT, ONNX, Hugging Face, OpenCV, scikit-learn, NCNN
Robotics & Vision: ROS/ROS2, Gazebo, MoveIt, PCL, Open3D, SLAM (ORB-SLAM, LIO-SAM), Motion Planning (OMPL, ARMTD)
ML Models & Algorithms: Transformers (BERT, GPT, Whisper), Vision Models (YOLO, SAM, DINO), Diffusion Models, Neural Rendering (NeRF, 3D Gaussian Splatting), Reinforcement Learning (PPO, SAC)
Tools & Platforms: Docker, Git, Conda, WandB, TensorBoard, CUDA Toolkit, AWS/GCP, Weights & Biases
3D/CAD Software: Fusion360, Blender, SolidWorks, MeshLab, CloudCompare, Mimics
Languages: Mandarin Chinese (Native), English (Fluent)