Building Robot Dexterity That Learns, Feels, and Adapts

Haozhi Qi
Amazon Frontier AI & Robotics (FAR)

General-purpose robots must be able to manipulate the physical world with the flexibility, precision, and robustness of human hands. While robots have achieved remarkable success in structured environments, **dexterous manipulation in unstructured settings** remains a critical bottleneck. Despite advancements in massive data collection [7], computation, and low-cost hardware [32], robots still struggle to perform intricate tasks that require rich physical contact, fine motor coordination, and robust adaptation.

My research aims to advance **generalizable robot dexterity**: the ability to interact with the physical world with human-level adaptability and robustness. In principle, anthropomorphic robot hands have the potential to provide the flexibility needed for such interactions [5, 6, 8, 22]. However, such flexibility also presents a fundamental trade-off. While multi-fingered hands enable complex contact-rich behaviors, they introduce significant algorithmic challenges, including data scarcity in high-dimensional action spaces, perception uncertainty during physical interaction, and difficulty transferring learned policies from simulation or humans to the real world.

To address these challenges, my research establishes multisensory learning and transfer as a foundation for robot dexterity. First, I develop methods for **scalable multisensory skill acquisition**, leveraging large-scale physical simulation to learn complex dexterous behaviors and transfer them to the real world [13, 14]. Second, I develop hardware and algorithms for **multisensory human-to-robot transfer**, including wearable tactile gloves [31] that capture contact-rich human demonstrations and Aria smart glasses that enable scalable egocentric data collection in everyday environments [11]. These systems reduce the embodiment gap and enable learning from human demonstrations with little robot data. Third, I develop methods for **real-world policy refinement**, using rich sensory feedback that is difficult to simulate [16]. *Together, these directions form a path toward robot dexterity that learns, feels, and adapts.*

A. Scalable Multisensory Skill Acquisition

A primary challenge in achieving generalizable dexterity is acquiring data for contact-rich interactions at scale. For instance, collecting teleoperation data for in-hand manipulation is significantly harder than for pick-and-place tasks with grippers [2]. Classical planning methods can produce precise contact-rich behaviors, but often rely on accurate contact or geometry models that limit generalization [10, 21, 26]. Learning-based approaches offer greater flexibility, but scaling them to dexterous manipulation remains difficult due to sample inefficiency on hardware and sim-to-real gaps in contact-rich

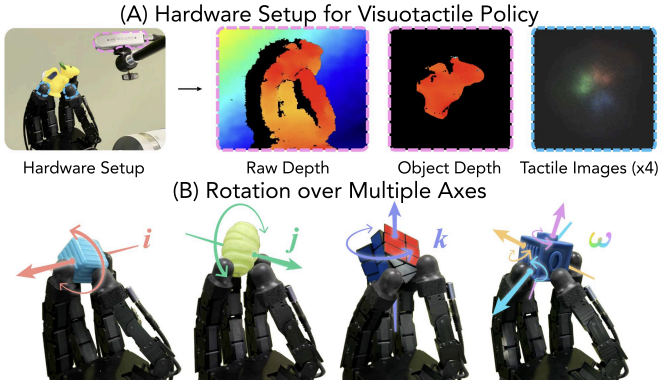

Fig. 1. **Scalable Skill Acquisition in Simulation**. Abstracting visual and tactile inputs into transferable representations bridges the sim-to-real gap, enabling in-hand rotation of diverse objects.

interaction [23]. My research addresses this challenge by using simulation as a scalable source of experience while abstracting raw sensory inputs into intermediate representations that transfer robustly to the real world.

My work introduces the *first framework for general in-hand object rotation conditioned explicitly on vision and touch* [14] (Figure 1 A, B). We use a two-stage learning pipeline. First, we train a policy using ground-truth 3D point clouds [25], showing that explicit 3D geometry improves performance on complex shapes. Second, we infer this shape from depth and tactile signals to bridge the sim-to-real gap. Because raw camera images and tactile deformations are difficult to simulate faithfully, we abstract *vision into object depth and touch into sparse contact locations*. During deployment, we extract object depth using pretrained vision models [18] and map tactile feedback to contact locations via fingertip sensors [19]. This representation reconstructs object shape while resolving perception ambiguities during contact-rich manipulation. We recently extended this framework to bimanual dexterous manipulation for a lid-twisting task [20]. This extension suggests that transferable visual and tactile abstractions can scale to higher-dimensional contact spaces.

B. Multisensory Human-to-Robot Transfer

Leveraging multifingered hands allows robots to learn from human demonstrations, one of the most scalable data sources [17, 28]. However, existing methods face two key limitations: scalability, due to cumbersome motion capture or VR setups, and sensory fidelity, due to limited tactile feedback. My research addresses these limitations through two portable systems and algorithms for **multisensory human-to-robot transfer**: `OSMO` for contact-rich tactile demonstrations and

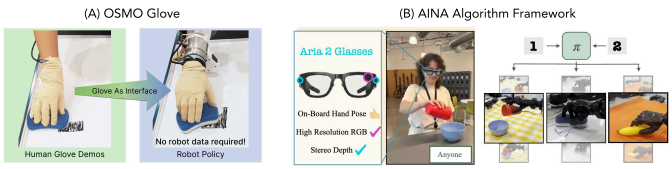

Fig. 2. **Human to Robot Skill Transfer**. (A) OSMO minimizes the visual and tactile embodiment gap with a shared wearable glove. (B) AINA uses lightweight smart glasses and 3D policy learning to scale human demonstration collection. Together, they enable robot policy learning from human demonstrations without physical robot data.

AINA for scalable egocentric demonstrations.

First, we introduce OSMO [31] (Figure 2 A), an open source wearable tactile glove. It features 12 **magnetic** three-axis tactile sensors across the fingertips and palm [3, 4], designed to work seamlessly with modern hand tracking [24]. By *putting the identical glove on both the human and the robot*, OSMO minimizes the visual and tactile embodiment gap and allows continuous shear and normal forces to be transferred directly. A robot trained exclusively on human demonstrations collected with OSMO successfully accomplishes contact-rich tasks. Second, we develop AINA [11] (Figure 2 B), an algorithmic framework for collecting human demonstrations with Aria smart glasses. Unlike bulky VR headsets, these lightweight glasses allow users to collect demonstrations in everyday environments. By lifting human videos to approximate 4D representations, AINA repurposes 3D policy learning to predict future hand keypoints and deploy the same policy on a dexterous robot hand. Operating in a shared space of 3D hand keypoints and 3D object point clouds reduces the human-robot domain gap, enabling **3D point based closed loop policies** trained exclusively on human demonstrations. Collectively, AINA scales visual data collection across diverse environments, whqile OSMO captures the multisensory contact signals needed for contact-rich tasks. Together, they enable generalizable policies for both vision-dominated and contact-rich manipulation without using physical robot data.

C. Closing the Sim2Real Gap via Multisensory Refinement

While the frameworks developed in the previous sections provide strong foundational policies, they cannot fully capture the physical uncertainties of real-world deployment. Complex tool use, such as *operating a screwdriver or fastening nuts and bolts* (Figure 3), is a critical milestone for robot manipulation. These tasks rely on rich tactile interactions that are difficult to simulate faithfully. While my first thrust abstracts tactile feedback into sparse contact locations, tool use requires continuous tactile signals to regulate force, detect slippage, and maintain stable contact. This motivates a secondary **multisensory refinement phase** for highly sensitive, contact-rich tasks.

To address this challenge, I propose DexScrew [16], a framework that bridges the sim-to-real gap through real-world tactile fine-tuning. The pipeline operates in three stages. First, we train reinforcement learning policies in simulation to learn robust finger gaits. Next, operators use these base policies as **skill primitives** for teleoperation, collecting real-world demonstrations with tactile and proprioceptive feedback.

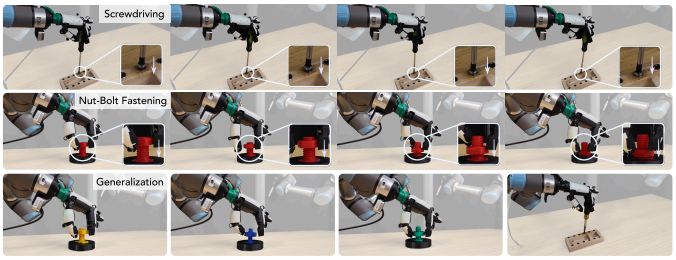

Fig. 3. **Tactile Guided Policy Refinement**. DexScrew bridges the sim-to-real gap for complex tool use by incorporating real-world tactile feedback. The closed-loop policy executes screwdriving and fastening with strong generalization across diverse geometries.

Finally, we train a behavior cloning policy that directly incorporates this multisensory feedback. This framework delegates complex finger coordination to simulation-trained policies while capturing real-world sensory observations during contact-rich manipulation. As a result, DexScrew generalizes to diverse nuts and screwdrivers, even under external perturbations. **Beyond tactile modalities**, real-world refinement could also address rigid-body dynamics mismatches in highly agile tasks, such as our learning-based method for **continuous pen spinning** [29] and **piano playing** [30].

D. Future Research

My future work will move from learning individual dexterous skills toward building continually improving systems for contact-rich manipulation:

Structured World Models and Continual Learning. *How can robots continuously update their understanding of an object's physical properties during active manipulation?* My work on NeuralFeels [27] showed that posed Signed Distance Fields (SDFs) unify multisensory perception. I will extend this toward structured world models that encode geometry, contact state, and physical properties from continuous tactile signals and RGB-D inputs. These models will enable model-based planning for long-horizon tasks and online adaptation to novel objects through continual learning [12].

Automating the Sim2Real Loop. *How can we systematically generate simulation environments that reflect complex real-world contact dynamics?* While DexScrew bridged the sim-to-real gap via teleoperation refinement, my future work will study how to automate this loop. I will investigate using Vision-Language Models (VLMs) and real-world demonstrations from the AINA pipeline to infer task structure, object affordances, and approximate physical parameters. These signals can guide targeted domain randomization and reward design, shifting sim-to-real transfer from manual tuning toward a scalable, data-driven process.

Composing Multisensory Skills via Language. *How can high-level semantic planners ground their decisions in low-level, continuous tactile feedback?* Long-horizon manipulation requires composing contact-rich skills while monitoring whether each skill is physically succeeding. I will build hierarchical systems where Large Language Models (LLMs) [1, 9] sequence libraries of low-level multisensory skill primitives [15], while tactile and proprioceptive feedback provide grounded signals for execution, failure detection, and recovery.

References

[1] Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Chuyuan Fu, Keerthana Gopalakrishnan, Karol Hausman, Alexander Herzog, Daniel Ho, Jasmine Hsu, Julian Ibarz, Brian Ichter, Alex Irpan, Eric Jang, Rosario Jauregui Ruano, Kyle Jeffrey, Sally Jesmonth, Nikhil J. Joshi, Ryan Julian, Dmitry Kalashnikov, Yuheng Kuang, Kuang-Huei Lee, Sergey Levine, Yao Lu, Linda Luu, Carolina Parada, Peter Pastor, Jornell Quiambao, Kanishka Rao, Jarek Rettinghouse, Diego Reyes, Pierre Sermanet, Nicolas Sievers, Clayton Tan, Alexander Toshev, Vincent Vanhoucke, Fei Xia, Ted Xiao, Peng Xu, Sichun Xu, Mengyuan Yan, and Andy Zeng. Do as i can, not as i say: Grounding language in robotic affordances. In *CoRL*, 2022.

[2] Sridhar Pandian Arunachalam, Sneha Silwal, Ben Evans, and Lerrel Pinto. Dexterous imitation made easy: A learning-based framework for efficient dexterous manipulation. In *ICRA*, 2023.

[3] Raunaq Bhirangi, Tess Hellebrekers, Carmel Majidi, and Abhinav Gupta. Reskin: versatile, replaceable, lasting tactile skins. In *CoRL*, 2021.

[4] Raunaq Bhirangi, Venkatesh Pattabiraman, Enes Erciyes, Yifeng Cao, Tess Hellebrekers, and Lerrel Pinto. Anyskin: Plug-and-play skin sensing for robotic touch. In *ICRA*, 2025.

[5] Antonio Bicchi. Hands for dexterous manipulation and robust grasping: A difficult road toward simplicity. *ICRA*, 2000.

[6] Aude Billard and Danica Kragic. Trends and challenges in robot manipulation. *Science*, 2019.

[7] Cheng Chi, Zhenjia Xu, Chuer Pan, Eric Cousineau, Benjamin Burchfiel, Siyuan Feng, Russ Tedrake, and Shuran Song. Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots. In *RSS*, 2024.

[8] Jinda Cui and Jeff Trinkle. Toward next-generation learned robot manipulation. *Science robotics*, 2021.

[9] Danny Driess, Fei Xia, Mehdi SM Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, Wenlong Huang, Yevgen Chebotar, Pierre Sermanet, Daniel Duckworth, Sergey Levine, Vincent Vanhoucke, Karol Hausman, Marc Toussaint, Klaus Greff, Andy Zeng, Igor Mordatch, and Pete Florence. Palm-E: An embodied multimodal language model. *arXiv:2303.03378*, 2023.

[10] Ronald Fearing. Implementing a force strategy for object re-orientation. In *ICRA*, 1986.

[11] Irmak Guzey, Haozhi Qi, Julen Urain, Changhao Wang, Jessica Yin, Krishna Bodduluri, Mike Lambeta, Lerrel Pinto, Akshara Rai, Jitendra Malik, Tingfan Wu, Akash Sharma, and Homanga Bharadhwaj. Dexterity from smart lenses: Multi-fingered robot manipulation with in-the-wild human demonstrations. In *ICRA*, 2026.

[12] Haozhi Qi, Xiaolong Wang, Deepak Pathak, Yi Ma, and Jitendra Malik. Learning long-term visual dynamics with region proposal interaction networks. In *ICLR*, 2021.

[13] Haozhi Qi, Ashish Kumar, Roberto Calandra, Yi Ma, and Jitendra Malik. In-hand object rotation via rapid motor adaptation. In *CoRL*, 2022.

[14] Haozhi Qi, Brent Yi, Sudharshan Suresh, Mike Lambeta, Yi Ma, Roberto Calandra, and Jitendra Malik. General In-Hand Object Rotation with Vision and Touch. In *Conference on Robot Learning (CoRL)*, 2023.

[15] Haozhi Qi, Brent Yi, Mike Lambeta, Yi Ma, Roberto Calandra, and Jitendra Malik. From simple to complex skills: The case of in-hand object reorientation. In *ICRA*, 2025.

[16] Elvis Hsieh, Wen-Han Hsieh, Yen-Jen Wang, Toru Lin, Jitendra Malik, Koushil Sreenath, and Haozhi Qi. Learning dexterous manipulation skills from imperfect simulations. In *ICRA*, 2026.

[17] Simar Kareer, Dhruv Patel, Ryan Punamiya, Pranay Mathur, Shuo Cheng, Chen Wang, Judy Hoffman, and Danfei Xu. Egomimic: Scaling imitation learning via egocentric video. In *ICRA*, 2025.

[18] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *ICCV*, 2023.

[19] Mike Lambeta, Tingfan Wu, Ali Sengul, Victoria Rose Most, Nolan Black, Kevin Sawyer, Romeo Mercado, Haozhi Qi, Alexander Sohn, Byron Taylor, Norb Tydingco, Gregg Kammerer, Dave Stroud, Jake Khatha, Kurt Jenkins, Kyle Most, Neal Stein, Ricardo Chavira, Thomas Craven-Bartle, Eric Sanchez, Yitian Ding, Jitendra Malik, and Roberto Calandra. Digitizing Touch with an Artificial Multimodal Fingertip. *arXiv preprint arXiv:2411.02479*, 2024.

[20] Toru Lin, Zhao-Heng Yin, Haozhi Qi, Pieter Abbeel, and Jitendra Malik. Twisting Lids Off with Two Hands. In *Conference on Robot Learning (CoRL)*, 2024.

[21] Andrew Morgan, Kaiyu Hang, Bowen Wen, Kostas E Bekris, and Aaron Dollar. Complex in-hand manipulation via compliance-enabled finger gaiting and multi-modal planning. *RA-L*, 2022.

[22] Allison M Okamura, Niels Smaby, and Mark R Cutkosky. An overview of dexterous manipulation. In *ICRA*, 2000.

[23] OpenAI, Marcin Andrychowicz, Bowen Baker, Maciek Chociej, Rafał Józefowicz, Bob McGrew, Jakub Pachocki, Arthur Petron, Matthias Plappert, Glenn Powell, Alex Ray, Jonas Schneider, Szymon Sidor, Josh Tobin, Peter Welinder, Lilian Weng, and Wojciech Zaremba. Learning dexterous in-hand manipulation. *IJRR*, 2019.

[24] Georgios Pavlakos, Dandan Shan, Ilija Radosavovic, Angjoo Kanazawa, David Fouhey, and Jitendra Malik. Reconstructing hands in 3d with transformers. In *CVPR*, 2024.

[25] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d

classification and segmentation. In *CVPR*, 2017.

[26] Daniela Rus. In-hand dexterous manipulation of piecewise-smooth 3-d objects. *IJRR*, 1999.

[27] Sudharshan Suresh, Haozhi Qi, Tingfan Wu, Taosha Fan, Luis Pineda, Mike Lambeta, Jitendra Malik, Mrinal Kalakrishnan, Roberto Calandra, Michael Kaess, Joseph Ortiz, and Mustafa Mukadam. NeuralFeels with Neural Fields: Visuo-Tactile Perception for In-Hand Manipulation. *Science Robotics*, 2024.

[28] Chen Wang, Haochen Shi, Weizhuo Wang, Ruohan Zhang, Li Fei-Fei, and C Karen Liu. Dexcap: Scalable and portable mocap data collection system for dexterous manipulation. In *RSS*, 2024.

[29] Jun Wang, Ying Yuan, Haichuan Che, Haozhi Qi, Yi Ma, Jitendra Malik, and Xiaolong Wang. Lessons from learning to spin "pens". In *CoRL*, 2024.

[30] Amber Xie, Haozhi Qi, and Dorsa Sadigh. Handelbot: Real-world piano playing via fast adaptation of dexterous robot policies. *arXiv:2603.12243*, 2026.

[31] Jessica Yin, Haozhi Qi, Youngsun Wi, Sayantan Kundu, Mike Lambeta, William Yang, Changhao Wang, Tingfan Wu, Jitendra Malik, and Tess Hellebrekers. Osmo: Open-source tactile glove for human-to-robot skill transfer. *RA-L*, 2026.

[32] Tony Z Zhao, Vikash Kumar, Sergey Levine, and Chelsea Finn. Learning fine-grained bimanual manipulation with low-cost hardware. In *RSS*, 2023.