Statement of Purpose – Jiahe Xu

My motivation in robotics comes from its potential to extend human capability—creating machines that perceive, adapt, and act in complex environments. What excites me most is the possibility of robots that learn from people naturally and generalize across tasks, hardware, and settings. Over the past years, I have pursued this vision by building the foundations for scalable robot learning through systems engineering, large-scale data pipelines, and experimental policy design. Now, I want to take the next step: to study how these foundations can be used to develop policies that are both scalable and reliable in a Ph.D. program.

At the CMU Air Lab, I gained my first experience in building complete robotic systems. I worked on real-time synchronization of heterogeneous sensors—cameras, LiDAR, IMU, radar, and GPS—and built calibration pipelines to keep them aligned. I also rewrote low-latency teleoperation code and experimented with using iPhones and iPads as lightweight sensing devices. These projects led to two publications, Flying Hand (2025 RSS) and Flying Calligrapher (2024 RAL), but more importantly, they taught me how to design systems that stay reliable in real-world operation.

In Prof. Katerina Fragkiadaki's lab, I shifted my focus from building platforms to building learning pipelines. I set up a 3D bimanual robot learning framework, starting with high-frequency depth value estimation for accurate perception, and then rewriting the Aloha robot's end-effector controller to achieve higher control frequency with less than 1 cm error. With this system in place, I benchmarked a wide range of VLA models on datasets such as RH20T, Open X-Embodiment, and BridgeData. I also created a cleaner and higher-frequency dataset of my own, which proved particularly valuable for training stable policies. Out of this effort came 3DFA (2026 ICRA, in submission), a policy that integrates flow matching with pretrained 3D scene representations to predict dense end-effector trajectories. Compared to earlier approaches, 3DFA not only improved accuracy in both unimanual and bimanual settings but also ran more than 30× faster, making real-time training and deployment practical.

After CMU, I co-founded PinocchioAI, where I extended these ideas into practice and confronted the broader limitations of current approaches. In real-world deployments, I found that VLA methods often lack robustness in precise 3D manipulation and struggle with embodiment transfer, as policies trained on one robot fail to generalize to new hardware. As the robot industry develops, new hardware and sensors enter the market every month; after a hardware improvement, much of the old data can no longer be used, which represents a huge waste of resources. To address this, my work aims to build agent-agnostic policies, allowing robots to reuse data and skills across hardware generations. I built an ecosystem where robots learn directly from humans, integrating multimodal pipelines for real-to-sim transfer and fine-tuned LLMs for interactive guidance. Most recently, I submitted OC3D (2026 ICRA, in submission), an object-centric diffusion framework for robot learning. These experiences reinforced my conviction that modular and standardized pipelines—combining perception, representation, human data, and real-to-sim—are essential for creating robots that can efficiently learn from human behavior with limited data and transfer across new tasks, environments, and hardware platforms.

Looking forward, I want my Ph.D. to focus on the open problem of how to best use standardized data and real-to-sim pipelines to train policies that are scalable, robust, and transferable. My long-term goal is to develop frameworks that enable robots to learn from humans efficiently, adapt across hardware generations, and operate reliably in diverse environments. More broadly, I hope my work will contribute to a future where robots are not just tools, but trusted partners that extend what people can achieve.