Open Robot Foundation Models: Development and Future Directions

This talk addressed the emerging paradigm of open robot foundation models and their role in shaping the future of physical AI and robotics. The speaker, a senior researcher with multiple academic and industrial affiliations in AI and robotics, framed the discussion around how recent advances in large-scale models, simulation, and data availability are fundamentally changing the way robots are designed, trained, and deployed.

The presentation began by situating robot foundation models within a broader historical context. While foundational studies in robotics and control date back decades, the convergence of large-scale computation, high-fidelity simulation, and global data sharing has enabled a qualitative shift in capability. Improvements in simulators, open-world environments, and parallel experimentation now allow robots to learn complex behaviors that were previously infeasible. This shift has fueled what the speaker referred to as a current “boom” in physical AI, where learning-based systems increasingly outperform traditional industrial robots in adaptability, even if they remain less precise from a classical engineering standpoint.

A central theme of the talk was the importance of predictive models in robotics. Rather than directly mapping perception to action, modern approaches emphasize learning predictive representations of the environment, enabling robots to reason over future states. These ideas, first explored several years ago, have gained renewed importance as they integrate naturally with foundation models trained on long-horizon video and multimodal data. The speaker highlighted demonstrations using long, unedited videos to capture rich task dynamics, arguing that such data is critical for robust generalization across tasks and environments.

The discussion then transitioned to the integration of whole-body control, perception, and learning, including recovery from noise, human teaching, and imitation learning. Advances in sensing and actuation now allow robots to be both powerful and compliant—for example, handling heavy objects while remaining safe for human interaction. This duality underscores the need for models that jointly reason about perception, control, and physical interaction, rather than treating them as isolated modules.

A significant portion of the talk focused on the organizational and infrastructural challenges of building open robot foundation models. These models require not only large datasets but also standardized tasks, shared benchmarks, and collaboration across academia, industry, and government. The speaker described ongoing efforts in Japan to establish open datasets, open-source software, and even open hardware, supported by industry partnerships and national strategies. Importantly, participation is not limited to traditional robotics companies, but also includes AI research groups and startups exploring teleoperation, avatar robotics, and large-scale human-in-the-loop systems.