11/21/2025

Keynote Lecture 1

By Jun Tani

This keynote lecture presented a comprehensive research program on goal-directed robot behavior grounded in active inference, with a particular emphasis on the emergence of executive control, visual attention, working memory, and compositional generalization across perception, action, and language. The work integrates cognitive inspiration with deep generative models to address out-of-distribution generalization in robotic manipulation.

The first part of the lecture focused on goal-directed action planning for object manipulation using a hierarchical generative model based on active inference. A robotic arm was trained to achieve visually specified goals by inferring sequences of visuoproprioceptive states rather than directly executing predefined actions. The architecture combined a higher-level executive LSTM with lower-level motor and vision prediction modules, incorporating attention mechanisms, pixel-wise masking, and a visual working memory. Through supervised training on approximately 300 manipulation trajectories, the robot learned to generate action plans that minimized variational free energy when primed with a goal image.

A central contribution of this work was the emergence of content-agnostic executive control. The visual working memory acted as a blackboard where object content and transformation functions were clearly separated. This structural separation enabled the robot to generalize to previously unseen object colors, such as purple, despite never encountering them during training. The model exhibited phenomena analogous to object permanence and event-level abstraction, where intermediate representations encoded the effects of actions rather than raw sensory data. These results highlighted how innate, differentiable structures—such as attention and masking—can facilitate efficient lifetime learning and robust out-of-distribution generalization.

The second part of the lecture addressed the development of compositionality between language and action. By extending the previous architecture with a language generation module, the system was trained to associate sentences (e.g., verb–noun combinations) with corresponding visuomotor trajectories. Importantly, the model was trained on only a subset of all possible verb–noun combinations and evaluated on unseen compositions. The results demonstrated that compositional generalization improved significantly as the size and diversity of the training vocabulary increased, even when the proportion of observed combinations remained constant. Internal latent representations revealed consistent clustering by action semantics across different object attributes, providing evidence of structured, compositional representations.

Finally, the lecture introduced a scalable, curiosity-driven reinforcement learning framework that integrated active inference principles. By rewarding actions that maximized expected free energy reduction, the robot engaged in exploratory, play-like behaviors that accelerated learning under sparse supervision. This developmental dynamic—where exploration increases model complexity and learning resolves it—enabled successful generalization across novel language commands, object shapes, and colors. Overall, the keynote demonstrated how combining active inference, executive control, and curiosity-driven learning offers a principled path toward scalable, cognitively inspired robot intelligence.