March 18 (Mon)

9:30-10:30

Prof. Jochen Triesch, Frankfurt Institute for Advanced Studies

"Developing Minds"

The human mind develops very differently from how most of today's artificial intelligence (AI) systems are trained. In particular, human infants understand the world around them by actively exploring it with their eyes and bodies rather than passively absorbing large training data bases. Understanding the computational principles that allow them to do so and modeling these principles in artificial systems is a grand challenge for Developmental Science and AI. In this talk, I will present our efforts to better understand milestones in infants' cognitive development through computational modeling and to develop AI systems that achieve these milestones in a similarly autonomous fashion. Central principles of our approach are that 1) developing infants actively shape their sensory input through their behavior and 2) they are intrinsically motivated to learn more abstract and efficiently compressed models of their experiences. I will show how by combining these principles it is possible to build, e.g., active vision systems that achieve early milestones of visual development including „understanding" the 3-D structure of space through active binocular vision, learning invariant object and category representations, or learning about the semantic relatedness of objects. I will conclude by describing how these models might be generalized to account for further milestones of cognitive development involving different sensory modalities.

10:30-11:30

Prof. Yasuo Kuniyoshi, The University of Tokyo

"From Embodiment to Super-Embodiment: An Approach to Open-Ended and Human Aligned Intelligence/Mind"

Embodiment is a key to solving the reliability and alignment issues in the current AI. This is because it imposes consistent constraints on the entire agent-environment interactions and accompanying information without specifying their actual contents, and the constraints are common to those with similar embodiment. The concept of embodiment should be generalized beyond mechanical properties of the body and

information structure of sensory signals, encompassing internal organs and metabolism, mental processes and inter-agent interactions, to “super-embodiment”. It can address sensibilities, values and morals toward artificial humanity, which will be critically important for the next generation AI.

11:30-12:30

Prof. Erhan Oztop, Ozyegin University

“Implications of Computational Resource Limitation on Cognition and Behavior”

To survive, animals need to perceive and act with the computational apparatus evolution has endowed them. Even the most evolved ones, such as humans, have to manage the computational resources they can devote to solving the problems they face. In decision theory, this concept is addressed under the concept of bounded rationality. This presentation will highlight the possible consequences of having to economize computational resources by introducing several computational modeling studies that follow this line of thought. In particular, ‘reuse of neural circuitry’ and ‘regulation of neural energy cost’ ideas will be exemplified in a cognitive neurorobotic context.

<u>March 19 (Tue)</u>

9:30-10:30

Prof. Stefano Nolfi, CNR-ISTC

“Integration and Transfer of Action and Language Knowledge in Learning Robots”

The integration of action, language knowledge and skills represents a key factor in human intelligence and is one of the most fascinating problems in science. In this talk I will review the evidence and insights collected by attempting to design robots capable of learning action and language skills. Specifically, I will focus on the contribution of foundational models and on the integration of methods that rely on passive observational learning and active embodied learning. Finally, I will discuss the potentials and limitations of alternative approaches.

10:30-11:30

Prof. Takashi Ikegami, The University of Tokyo

“Emerging Autonomy in Robots Powered by Large Language Models”

ChatGPT, which emerged in 2023, and the subsequent LLM (Large Language Model) have completely transformed programming for robots. Moreover, they aim to endow robots and swarms of robots (e.g. 2,3) with unprecedented autonomy, self-awareness, higher-level conversations, and meta-behaviors. Here, we report on the autonomy and collective intelligence generated in real robots and agent collectives driven by LLMs (e.g. 1) .

references:

1. Takahide Yoshida, Atsushi Masumori and Takashi Ikegami: **From Text to Motion: Grounding GPT-4 in a Humanoid Robot "Alter3"** https://arxiv.org/abs/2312.06571
2. Ryosuke Takata, Yujin Tang, Yingtao Tian, Norihiro Maruyama, Hiroki Kojima, Takashi Ikegami. **Evolving Collective AI: Simulation of Ants Communicating via Chemicals**, Proceedings of the 2023 Artificial Life Conference
3. Norihiro Maruyama, Michael Crosscombe, Shigeto Dobata, Takashi Ikegami, Emergence of **Differentiation of Deterministic/Stochastic Behavior in Ants’ Collectives**, Proceedings of the 2023 Artificial Life Conference.

11:30-12:30

Prof. Tom Froese, OIST

“LLMs, conversational memory, and the minimal self”

AI systems based on LLMs have been spectacularly successful and they are sure to go down in history as a game-changing technology. They have convincingly mastered what the famously AI-skeptical philosopher Dreyfus had called the “common sense knowledge problem.” In fact, these conversational systems have been so effective that old debates regarding the prospects of machine consciousness have been reignited in both academic and public domains. From the perspective of Dreyfus’ position, and that of embodied cognition more generally, it may seem intuitive to continue to insist on rejecting these technological breakthroughs as nothing but scaled-up forms of computational-statistical modeling. This resistance derives from the fact that LLM-based systems have essentially been dis-embodied and un-situated, although this situation is quickly changing. Moreover, there has been a notable transformation of cognitive architecture with the recent advent of recursive implementations of conversational memory in LLMs, whereby they are

prompted at regular intervals to contextually carry forward only the gist of the past conversation. In this talk I will analyze this recursive sense-making capacity of LLMs in the context of the tripartite structure of time consciousness, which according to the phenomenological tradition is the structural basis for a minimal self.

14:00-15:00

Prof. Kohei Nakajima, The University of Tokyo

“Physical reservoir computing and beyond”

Physics reservoir computing (PRC) is an approach to implementing reservoir computing in physical systems. While the advancement of PRC is remarkable in recent days, it remains at the level of principle demonstration in many cases. It is expected that as the scalability of the systems progresses in the future, their performance will reach to similar levels of the state-of-the art machine learning systems. In this presentation, we introduce methods to improve performance even in current small-scale systems. This involves coupling small physical reservoirs, which is called deep reservoir computing architecture, along with a new learning method called augmented direct feedback alignment (aDFA). In general, obtaining precise models for physical systems is difficult. Therefore, calculating gradients to optimize the weights coupling the physical systems for a task becomes even more challenging. Here, we introduce the method of aDFA, which allows weight learning without obtaining gradients, and demonstrate its effectiveness in various settings and implementation experiments. Finally, we discuss the scope and implications of deep physical reservoir computing.