# AVENUES OF INTELLIGENCE WORKSHOP

# HEIDELBERG UNIVERSITY

# AUG 29 & 30, 2024

| **Day I** | |
|---|---|
| 09:30-09:45 | *Welcome And Introduction* I *Dr. Antonio Bikić* |
| 09:45-10:30 | *1. Computation and control in animal brains and neuromorphic hardware*<br><br>**Prof. Dr. Yulia Sandamirskaya**<br>**Cognitive Computing, Zurich University of Applied Sciences** |
| | **Abstract:** Biological brains have evolved to control movement and they excel in this task. Integrating input from distal (visual, auditory) and proximal (touch, proprioception, motor sense) sensors, even simple animals can navigate to goals in the environment, form memories and motion plans, and control motion when unexpected disturbances happen, learning and getting better as they do so. To the contrary, today's deep neural networks – inspired by the early understanding of the brain – can perform pattern classification and function approximation with an unprecedented performance, but they are extremely inefficient in computing time and energy footprint, amount of data needed to learn, and inability to deal quickly with novel situations. What can we learn from modern neuroscience for the next generation of neural-networks based AI systems that would be capable to control movement in the real world? What are the required building blocks and computing principles? What is the difference between these principles and today's computing architectures, including deep learning? I will shed some light on these questions in the talk, sketching a novel computing framework, which marries computation with control. |
| 10:30-11:15 | *2. Vector Epistemology: Reasons as Vectors*<br><br>**Prof. Dr. Dr. Hannes Leitgeb**<br>**Philosophy/Mathematics, Munich Center for Mathematical Philosophy, LMU Munich** |
| | The talk presents an axiomatic theory of epistemic reasons (reasons for belief), argues that the theory might serve as a formal epistemology for the dynamics of artificial neural networks, and thereby builds a bridge between philosophy and AI. The upshot will be: vectors and vector operations in neural networks correspond to epistemic reasons and epistemic operations on such reasons, respectively. |
| 11:15-11:30 Uhr | *Coffee Break* |
| 11:30-12:15 | *3. The Reasons of AI Systems*<br><br>**Prof. Dr. Eva Schmidt**<br>**Philosophy, Technical University Dortmund** |

| Day I | |
|---|---|
| | **Abstract:** This talk connects the fields of philosophy of action and of explainable artificial intelligence (AI). We investigate whether it can ever be appropriate to explain the outputs of AI systems by appeal to practical reasons and reasoning of these systems. We argue that this can indeed be fitting. To this end, we first present an argument that starts from the premise that we use AI systems because they work so well for us and then defend our claim against four objections. (Co-Authors: Kevin Baum, Maximilian Schlüter, and Timo Speith) |
| 12:15-13:45 | *Lunch* |
| 13:45-14:30 | *4. Black Boxes as Explainers*<br><br>**Prof. Dr. Michael Strevens**<br>**Philosophy, New York University** |
| | **Abstract:** "Black-box" machine learning models of complex processes such as protein folding and weather systems can be powerful predictive devices, but they do not seem to offer any explanatory insight—any insight into the reasons that the phenomena they predict so well occur. This paper investigates ways in which we might derive understanding from machine learning "oracles", by regarding them as in part, if not wholly, engaged in a search for higher-level or (in a certain sense) emergent causal connections. Extracting a machine learning model's "knowledge" of such connections may not be easy, but there may be understanding there to be mined. |
| 14:30-15:15 | *5. Intrinsic Complexity of Inner Interpretability*<br><br>**Dr. Federico Adolfi**<br>**Computational Cognitive Science/Theoretical Computer Science; University of Bristol & ESI, Max-Planck Society** |
| | **Abstract:** Emerging folklore in AI suggests that combining inner-interpretability techniques with the expressivity, observability and perturbability of artificial neural networks will give us a fundamentally novel and more efficient methodology for scientific discovery and theory-building. Although there is optimism among cognitive scientists and engineers, we currently lack a principled understanding of what interpretability algorithms could offer or how they could plausibly deliver. I will preview results of computational complexity-theoretic modeling of interpretability tasks that capture various notions of mechanistic understanding of neural networks. After bringing these results to bear on real-world problems of interest for AI, Neuroscience, and Psychology, I will draw out the broader implications for engineering, the cognitive sciences, and the philosophy of AI-fueled discovery. |
| 15:15-15:30 | *Closing Plenary Session* I *Dr. Antonio Bikić* |

| **Day II** | |
|---|---|
| 09:00-09:15 | *Welcome* I *Dr. Antonio Bikić* |
| 09:15-10:00 | *6. Basal Cognition and the Case for Medium Dependence*<br><br>**Dr. Mazviita Chirimuuta**<br>**Neuroscience/Philosophy, University of Edinburgh** |
| | **Abstract:** In this talk I argue that new research in the field of basal cognition puts pressure on the assumption that information processing in biological systems can be medium independent, as assumed in the classical computational theory of mind. A key finding is that information processing occurs within organisms at multiple nested scales (McMillen and Levin 2024). This means, for example, that the functionality of the whole brain and nervous system depends on there being a smaller-scale version of this functionality at the level of circuits and neurons. For a system to implement this fractal-like organisation, its material components must reach a threshold of organisational complexity -- they could not be blobs of Swiss cheese! Hence the case for medium dependence. |
| 10:00-10:45 | 7. *Neural Geometry*<br><br>**Dr. Nico Scherf**<br>**Computer Science, Max Planck Institute für Human Cognitive and Brain Sciences** |
| 10:45-11:00 | *Coffee Break* |
| 11:00-11:45 | *8. The differences between AI and human object recognition - scale is not what we need*<br><br>**Dr. Marin Dujmović**<br>**Psychology, University of Bristol** |

| Day II | |
|---|---|
| | **Abstract:** The AI boom, which has taken the general public by storm recently, had started for researchers in 2012 when a convolutional deep neural network, AlexNet, won the ImageNet large scale visual recognition competition by a wide margin. Since then, deep neural networks have become more complex following the massive increase in computational resources. Following suit, datasets used to train them have also ballooned. For some researchers “scale is all you need” – meaning, huge naturalistic datasets and ever more parameters to train and tune will bring upon the I in AI. It has not only been expected to increase performance of networks, but lead to good alignment with the human brain. In this talk I will present some of the work done during the “Generalization in mind and machine” project which demonstrates clear differences in how deep neural networks and humans achieve object recognition and how scale may not be what we need. A sense of objecthood, shape bias, gestalt effects and a host of other basic visual phenomena do not emerge from simply increasing scale. Indeed, scale and computational power may be leading to very unhuman-like systems with all the implications that entails. |
| 11:45-12:30 | *9. Learning and Perception of Actionable Representations for Embodied AI*<br><br>**Prof. Dr. Jörg Stückler**<br>**Computer Science, University of Augsburg** |
| | **Abstract:** Intelligent robots require the abilities to learn representations of their action capabilities in the environment for reasoning and control and to align their internal representations with sensory information. In this talk, I will present my recent research on learning and perception of such actionable representations. I will detail self-supervised and physics-informed approaches that learn action-conditional dynamics models of objects and the robot. I will also provide an overview of current trends in embodied vision and artificial intelligence. |
| 12:30-12:45 | *Closing remarks* I *Dr. Antonio Bikić* |