Hybrid Plasticity: Adaptive, Brain-Like Artificial Intelligence via Prefrontal Cortex Inspired Meta-Learning

Kalahasty, Rohan (School: Thomas Jefferson High School for Science and Technology)

Current artificial intelligence (AI) systems can learn arbitrarily complex patterns but struggle to generalize and adapt to new situations, making them unsuitable for real-world deployment. AI is also highly inefficient, with GPT-3 outputting upwards of 1,100,000 pounds of CO_2 so far. Inspired by the workings of the core of human intelligence – the prefrontal cortex – I propose Hybrid Plasticity, which allows AI to dynamically rewire itself given situation-specific inputs after training. During training, synapse-specific HP coefficients are learned, allowing network weights to use environmental input to guide convergence to optimal weights as governed by an attractor in the weight-phase-space. In OpenAI's gym, agents imbued with HP perform on par or better than state-of-the-art reinforcement learning methods, including OpenAI's proximal policy optimization and DeepMind's neural attention, while also being able to adapt to damage and having more stable, robust performance. In recurrent neural networks completing complex memory tasks, HP greatly increases adaptability and efficiency while causing close mimicry of neurobiological phenomena within network connectivity; networks self-organize into small world networks – structures posited to cause consciousness, exhibit inhibitory and excitatory neuron clustering, develop strong inhibitory autapses, and display high levels of parallel processing and cognitive integration. This provides evidence for two theories, the Synaptic Theory of Memory and the Global Workspace Theory, indicating that HP can be used to study the brain in-silico. Experimentation demonstrates that HP can reduce the carbon footprint of AI by ~20% and allow AI to adapt to unforeseen circumstances, effectively making a crucial step towards strong AI.

Awards Won:

Third Award of $1,000