IRGF Final Report

Eric Chalmers - #103309

Project Title: Building a Brain-Inspired Artificial Intelligence to Model Human Learning and Disorders

1.1 Project Summary

Creating Artificial Intelligence (AI) algorithms that learn in more human-like ways could allow us to imbue AI with more human-like abilities. Interestingly, it can also allow us to use AI to model and explore dysfunctions of human learning, such as mood disorders. This project drew inspiration from the brain to produce AI programs that demonstrate flexible, human-like planning, and can also experience depression.

1.2 Project Outcomes and Impacts

Animals can navigate through complex environments with amazing flexibility and efficiency: they forage over large areas, quickly learning rewarding behavior and changing their plans when necessary. This type of flexibility is difficult for AI. Some insight into how the brain achieves such flexibility can be found in the hippocampus - a brain structure involved in navigation, learning, and memory. We explored experimental results from neuroscience research and highlighted three principles suggested by the way the hippocampus encodes and processes information: 1) learning is hierarchical, 2) learning a hierarchical world-model is intrinsically valuable, 3) and action planning occurs as a downstream process separate from learning. By writing new AI algorithms based on these principles, we demonstrated that an AI agent could learn hierarchical models of a complex problem, and that this allows computationally efficient planning to reach arbitrary goals. Our results help to clarify different interpretations of some spatial navigation studies in rodents and present some implications for future studies of both machine and biological intelligence. In parallel, we explored another benefit of recreating human-like learning mechanisms in AI: the ability to create AI models of learning dysfunctions, including mood disorders. We impaired an artificial neural network to simulate the loss of interneuron connectedness that has been observed in depressed brains. This alteration caused the AI agent to demonstrate a wide variety of depression-like behaviors. Surprisingly, modelling other longstanding conceptual models of depression did not produce the same range of depression-like behaviors, indicating good face validity for our AI model of depression. Our model offers some insight into depression's underlying mechanisms, and offers a way to reconcile new and older theories of depression pathology. The student researchers supported by this IRGF award provided brilliant contributions to these projects and gained valuable experience. They have now moved onto other research and leadership roles.