The Weizmann Institute of Science
Faculty of Mathematics and Computer Science

Vision and AI

on Thursday, Jun 13, 2024
at 16:00

ZOOM

Zoom Only

James DiCarlo
MIT

will speak on

Today's "AI"-derived engineer-level models of the neural mechanisms of primate object perception, and tests of their application in non-invasive modulation of high level brain states

Abstract:
The human species is on a great scientific quest — to understand the neural mechanisms of human (primate) intelligence. Recent progress in multiple subfields of brain research suggests that key next steps in this quest will result from building real-world capable, systems-level network models that aim to abstract, emulate and explain the primate neural mechanisms underlying natural intelligent behavior. In this talk, I will briefly outline the story of how neuroscience, cognitive science and computer science ("AI") converged to create specific, image-computable, deep neural network models intended to appropriately abstract, emulate and explain the mechanisms of primate core visual object identification and categorization. Based on a large body of primate neurophysiological and behavioral data, some of these network models are currently the leading (i.e. most accurate) scientific theories of the internal mechanisms of the primate ventral visual stream and how those mechanisms support the ability of humans and other primates to rapidly and accurately infer latent world content (e.g. object identity, position, pose, etc.) from the set of pixels (image) received under typical (brief) natural viewing.

While still far from complete, because these leading neuroscientific models are fully observable and machine-executable, they offer predictive and potential application power that the field's prior conceptual models did not.   I will describe two recent examples from our team. First, I will show that — relative to the gold standard of primate brains and minds — the leading models are both similarly sensitive to and similarly robust to adversarial attack at both their neural levels and their behavioral levels. Second, I will describe initial empirical tests of the closely related possibility of using such models to design spatial patterns of light energy on the retina (i.e.

in the primate brain. Consistent with model predictions, these tests reveal surprisingly strong and precise neural population effects. Besides being a tool for neuroscience, we see this as an exciting new application avenue of potential human clinical benefit.