

Part-whole structure in visual foundational models

Foundation models provide a model system that (arguably) exhibits visual intelligence, is fully accessible to intervention and readout at all levels, and can communicate about its own internal states using natural language. Not only are such models an obvious experimental subject for visual neuroscience, they are also scientifically interesting objects of study in their own right. In this talk I'll describe early-stage work aimed at probing how visual information is represented and transformed in these models, drawing on ideas from visual psychology, systems neuroscience, and vision theory. I will focus on two particularly fascinating aspects of visual function: mental imagery, and part-whole structure. To conclude, I will outline why understanding vision in foundation models is naturally a neuroscience problem—and to invite discussion about what the right questions and tools should be.