How Will the AI Genie Behave

We are witnessing the increasing ubiquity of AI, with accelerating deployment in a wide variety of task domains, entrepreneurial activity across industries, and embedding in the core processes of digital platforms and firms. Alongside the increasing ubiquity of AI, we are encountering broad unintended consequences, both beneficial and harmful. These unintended consequences are arising from unexpected behaviors and downstream effects that emerge through interactions of AI agents with the environment and other agents. To be able to control the actions of AI in ways that harvest its benefits while minimizing harm, we need to develop an integrated understanding of AI behavior in the environments in which it is deployed. I will offer a perspective on how such an integrated understanding can be developed through research that investigates thorny issues related to how AI agents acquire their behavior, how the behavior is triggered and enacted, the interaction dynamics of AI agents with other agents that lead to unintended consequences, and the influence of policy decisions on how AI systems evolve.