Benjamin Manning
PhD Candidate, Management Science & Information Technology
MIT, Sloan School of Management

Thursday, October 27, 2026
When: 12:00 PM to 1:20 PM
Where: 3730 Walnut Street, Suite 700

GENERAL SOCIAL AGENTS

Useful social science theories predict behavior across settings. However, applying a theory to make predictions in new settings is challenging: rarely can it be done without ad hoc modifications to account for setting-specific factors. We argue that AI agents put in simulations of those novel settings offer an alternative for applying theory, requiring minimal or no modifications. We present an approach for building such "general" agents that use theory-grounded natural language instructions, existing empirical data, and knowledge acquired by the underlying AI during training. To demonstrate the approach in settings where no data from that data-generating process exists—as is often the case in applied prediction problems—we design a heterogeneous population of 883,200 novel games. AI agents are constructed using human data from a small set of conceptually related but structurally distinct "seed" games. In preregistered experiments, on average, agents predict initial human play in a random sample of 1,500 games from the population better than (i) various economic models based on level-k reasoning, (ii) game-theoretic equilibria, and (iii) out-of-the-box agents. For a small set of separate novel games, these simulations predict responses from a new sample of human subjects better even than the most plausibly relevant published human data.