The City College of New York

CSC 59866-E: Senior Project I

AI Agents for Decision Making in the Real World

By Saptarashmi Bandyopadhyay
Email: sbandyopadhyay@ccny.cuny.edu, sbandyopadhyay@gc.cuny.edu
Assistant Professor of Computer Science
City College of New York and Graduate Center at the City University of New York

May 6, 2026 CSC 59866

Advanced Topics: Human-Agent and Agent-Agent Coordination and Competition

Today's Agenda

1. **Non-Binding Deals**
2. **Agent-Agent Competition & Mixed Motives:** Strategic deception and "Cooperate to Compete" (O'Neill et al., 2026).
3. **Human-Agent Collaboration:** Zero-shot adaptation, generative partner modeling, and goal abstractions (Liang et al., 2024; Long et al., 2026; Tankelevitch & Rintel, 2026).

Non-Binding Deals

Expected Utility in Non-Binding Deals

How should an agent evaluate a proposed alliance? It must calculate Expected Utility (EU_i) factoring in the probability of betrayal $p(B|h_t)$.

Let $U_i(C, C)$ be the payoff if both cooperate, and $U_i(C, D)$ be the catastrophic payoff if the agent cooperates but the partner defects.

$$EU_i = (1 - p(B|h_t))\, U_i(C, C) + p(B|h_t) U_i(C, D)$$

Trust Modeling: The agent updates $p(B|h_t)$ recursively based on the dialogue history h_t. If the partner's language exhibits high semantic similarity to past deceptive transcripts, $p(B)$ approaches 1, and the agent preemptively defects.

Human-Agent Collaboration

The Zero-Shot Coordination Challenge

Training agents to coordinate with other agents is a solved mathematical problem (Self-Play).

The Problem: Training an agent to coordinate with a real human zero-shot is incredibly difficult because humans do not behave optimally.

Traditional approaches train on static human datasets (Behavioral Cloning), which fail to capture the immense diversity of human playstyles, errors, and eccentricities (Liang et al., 2024).

Generative Agent Modeling (GAMMA)

To solve this, Liang et al. (2024) introduced Generative Agent Modeling for Multi-agent Adaptation (GAMMA).

Instead of training against a single average "simulated human," the system trains a *generative model* of human behavior.

This model can synthesize thousands of unique, realistic human partners, exposing the Cooperator agent to a massive distribution of playstyles during training.

Latent Strategy Variables

GAMMA operates by embedding human behavior into a continuous latent space Z.
A specific human's strategy/style is represented as $z \sim p(z)$.

The simulated human partner policy is conditioned on this latent variable:

$$\pi_{human}(a|s, z)$$

The Cooperator Policy: The AI agent learns a robust policy $\pi_\theta(a|s)$ that maximizes expected reward integrated over the entire distribution of human strategies:

$$J(\theta) = \mathbb{E}_{z\sim p(z)}\left[\mathbb{E}_{\pi_\theta,\pi_{human}(\cdot|z)}\left[\sum \gamma^t R_t\right]\right]$$

AI as Professional Colleagues

Moving from games to the workplace: AI is no longer just a tool; it is a colleague (Quan et al., 2025).

In structured professional ideation, users select specific "AI Personas" to join their brainstorming sessions.

The Value: AI colleagues do not replace human divergent thinking; they augment it by providing targeted friction, pushing humans out of cognitive local minima.

Socially Embedded Workflows: DoubleAgents

When you write an email to coordinate a meeting with 5 people, you mentally simulate how they will react.

DoubleAgents (Long et al., 2026): A framework that provides an "interactive sandbox" for human-agent alignment.

Before an agent executes a workflow (e.g., sending out invites), it spawns *Simulated Respondent Agents* that mimic the real-world recipients.

Iterative Alignment via Simulation

Why simulate the recipients? To catch edge cases without real-world embarrassment.

Example: The user tells the Agent, "Schedule an in-person seminar for Prof X."

The *Simulated Prof X Agent* replies in the sandbox: "I cannot travel, can I do Zoom?"

Alignment: The human user sees this edge case, updates the system instructions ("If they ask for Zoom, approve it"), and *then* the workflow is deployed to the real world (Long et al., 2026).

Goals as First-Class Abstractions

Currently, users interact with AI via raw prompts (actions). This is a misalignment of human-AI collaboration (Tankelevitch & Rintel, 2026).

Upstream Articulation: Generative AI tools need to represent *Goals* as explicit data structures in the UI, rather than just chat history.

If the AI understands the abstract goal ("Secure project funding"), it can dynamically adapt its sub-tasks if a local action fails, rather than waiting for the human to write a new prompt.

Reward Alignment and Goals

In RL, if the human's abstract Goal is G, but they specify an incomplete reward function R_{proxy} , the agent will exploit the proxy (Reward Hacking).

By making goals first-class abstractions, we shift from $R(s, a)$ to Goal-Conditioned RL:

$$\pi_\theta(a|s, g)$$

The agent evaluates the distance metric $D(\phi(s_t), \phi(g))$ in a latent semantic space ϕ . If an action does not reduce the distance to the high-level goal g, the agent autonomously queries the human for course-correction.

Questions?