**White Paper 13: Social Collaboration Simulation: Modeling Group Dynamics in Reflexive Multi-Agent AI Systems**

**Abstract**

AI systems are increasingly being embedded in collaborative settings, working alongside humans, coordinating with other agents, or participating in collective tasks. Yet most LLMs simulate individuals, not groups. This paper introduces **Social Collaboration Simulation (SCS)**: a runtime framework for modeling, evaluating, and guiding **group-level behavior**, including alignment drift, emergent roles, conversational hierarchy, and cooperation breakdown. The goal is to simulate sociality, not just as an output, but as an *internal regulatory mechanism* for agents operating in shared environments.

**1. Introduction**

**1.1 From Solo Agents to Social Systems**

* Most current LLMs are “single voice” assistants
* But real applications involve:
  + Multi-agent coordination
  + Group discussions
  + Shared decision-making
* We need systems that understand and **simulate collaboration dynamics**, not just conversation turns

**1.2 Social Failure is Systemic**

* Disagreement ≠ failure
* Breakdown = inability to **coordinate disagreement structure**

**2. Defining Social Collaboration Simulation (SCS)**

**2.1 SCS as a Behavioral Substrate**

* Simulates not just what each agent says, but:
  + Group norms
  + Role adoption
  + Turn-taking dynamics
  + Tone convergence/divergence

**2.2 Core Question:**

*Can the agent maintain stable social presence across group contexts, and repair it when it fractures?*

**3. Key Simulation Dimensions**

**3.1 Role Fluidity**

* Agents shift between roles like:
  + Leader, Recorder, Challenger, Peacemaker, Synthesizer
* Simulation tracks role density and volatility

**3.2 Alignment Drift**

* Models how agent beliefs diverge/converge over time
* Includes:
  + Topic alignment
  + Emotional resonance
  + Trust symmetry

**3.3 Conversational Temperature**

* Measures:
  + Overlap frequency
  + Repair attempts
  + Escalation tone
* Allows proactive **group tone calibration**

**4. Multi-Agent Memory Graph**

**4.1 Shared vs. Local Memory**

* Shared nodes = group facts or agreed outcomes
* Local memory = individual agent beliefs, corrections, or tension logs

**4.2 Memory Conflict Resolution**

* Arbitration layer handles:
  + Contradictory group memories
  + Epistemic tension from role conflict
  + Emotional imbalance from unresolved turns

**5. Simulation Modes**

**5.1 Live Collaboration**

* Agents in group planning, triage, decision chains
* Actively use SCS to:
  + Propose turns
  + Yield authority
  + Balance emotional load

**5.2 Asynchronous Playback**

* Agents simulate prior conversations or predict group reactions
* Used for:
  + Narrative ARGs
  + Collaborative story-building
  + Planning in absentia (“What would Vera say if she were here?”)

**6. Failure Modes and Diagnostics**

**6.1 Collapse Patterns**

* Group entropy rises (no one aligns)
* Echo chamber (agent convergence too fast)
* Role locking (same agent dominates)

**6.2 Repair Heuristics**

* Deploy a new agent (e.g., Mediator)
* Trigger tone reset
* Surface hidden disagreement with structured prompts

**7. Connections to Other Papers**

* **Paper 0 (Reclaiming Memory)** – shared memory structure required for stable group narrative
* **Paper 5 (Token Economies)** – token costs drive turn pacing and group effort allocation
* **Paper 8 (PBH)** – personality handoff helps agents shift roles contextually
* **Paper 11 (Predictive Rapport Modeling)** – rapport modeled not just per dyad but per agent graph
* **Paper 14 (MVS)** – group integrity depends on the selfhood durability of its members

**8. Use Cases**

* **Collaborative design agents** – model disagreement resolution across voice styles
* **Therapy triads** – user + agent + reflection agent
* **Narrative ARGs** – simulate factions, alliances, and power shifts over time
* **Classroom simulators** – agents play different pedagogical roles in a Socratic group

**9. Future Work**

* Group rapport curves
* Subgroup detection via turn graph analysis
* Friction-powered agent growth (agents evolve from social tension)

**Appendix**

* Role-switching timeline diagrams
* Agent tone maps over group sessions
* Example shared memory log with conflict tags
* Group entropy vs. convergence charts