How Work Really Gets Done: Activity Theory and Persona-Based Social Reasoning for AI Agents

Lucy Lu
University of Waterloo, David R. Cheriton School of Computer Science
Waterloo, Canada
l594li@uwaterloo.ca

Lucas Gomez Tobon
University of Waterloo, David R. Cheriton School of Computer Science
Waterloo, Canada
lgomezto@uwaterloo.ca

Edith Law
University of Waterloo, David R. Cheriton School of Computer Science
Waterloo, Canada
edith.law@uwaterloo.ca

ABSTRACT

As AI agents move toward proactive roles in human organizations, a key challenge is enabling them to navigate complex, evolving social environments. To act competently, agents must understand not only explicit tasks but also the tacit norms, relationships, power dynamics, and informal rules that shape organizational life. We propose an approach for building socially competent AI agents through organizational activity modeling and persona-based reasoning. Drawing on Activity Theory, our system generates dynamic representations of tasks and social structures from communication streams (emails, chats, calendars), creating persistent organizational memory that informs agent behavior. Using retrieval-augmented generation, agents access relevant social knowledge in real time, supporting adaptation to changing contexts. We outline a three-stage research plan to evaluate this framework, from synthetic crisis scenarios to real-world deployments in asset management firms. Our goal is to advance the design of AI agents capable of integrating into, and evolving with, human organizational environments.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; • **Computing methodologies** → **Multi-agent planning**.

KEYWORDS

AI agents, organizational social knowledge, activity theory, persona modeling, human-AI collaboration, social reasoning, organizational integration, situated AI

ACM Reference Format:
Lucy Lu, Lucas Gomez Tobon, and Edith Law. 2025. How Work Really Gets Done: Activity Theory and Persona-Based Social Reasoning for AI Agents. In *Proceedings of Social Agentics Workshop: Situating Agentic AI Within Social & Organizational Contexts (Social Agentics @ COMPASS 2025).* ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

Social Agentics @ COMPASS 2025, July 22–25, 2025, Toronto, ON, Canada

ACM ISBN 978-x-xxxx-xxxx-x/25/07
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

As AI agents transition from reactive tools to proactive, autonomous participants within human organizations, their ability to act competently depends on more than technical proficiency in task execution. Real-world organizational environments are structured by complex, evolving networks of roles, relationships, tacit norms, and informal practices that fundamentally shape how work gets done. Yet today's AI agents lack mechanisms to perceive, model, or reason about these layered social contexts. Current architectures, largely grounded in large language model (LLM) pretraining or narrow task-specific fine-tuning, do not equip agents to adapt to the implicit protocols, hierarchies, and coordination practices that govern human collaboration. In particular, they lack explicit representations to model, update, and reason about social context. This gap fundamentally limits the agent's capacity to participate meaningfully in human teams, impairing its ability to identify relevant collaborators, select appropriate communication channels, understand authority structures, or anticipate social dynamics around task flow. Advancing truly situated, socially capable agentic systems requires fundamental progress in how AI agents acquire, represent, and utilize organizational social knowledge.

Imagine you are a senior project manager at a large organization, working with an AI-based executive assistant that helps coordinate meetings and team workflows. You ask the agent: "Please find a good time this week for an in-depth project milestone review with the full team". On the surface, this would appear to an AI agent to be a simple scheduling task. But successfully navigating it requires far more than comparing calendar availabilities. Your team is hybrid, spanning multiple time zones and departments—including engineering, marketing, and legal—each with its own work rhythms and communication practices. Some team members are senior leadership, others are junior staff. Certain colleagues have a history of interpersonal tension; others are navigating high workload stress. Unspoken norms discourage scheduling late in the week, while department practices require legal pre-briefing before risk discussions. Communication preferences vary across formal calendar systems and informal messaging platforms. Even deciding who must review and approve the agenda reflects invisible hierarchies and organizational politics, knowledge not captured in any database. What seems like a routine task quickly reveals layers of social complexity. For an AI agent to perform competently in this context, it must develop an evolving, situated understanding of relationships, roles, informal norms, and contextual sensitivities, capabilities that today's systems, built on static pretraining, do not yet possess.

We propose a systematic approach to social competence through organizational activity modeling and persona-based reasoning with

retrieval-augmented generation (RAG) [13]. Drawing from Activity Theory's framework [7, 8] for understanding goal-directed collaborative behavior, our system constructs dynamic representations of organizational tasks that capture functional requirements alongside social structures, power dynamics, and communication patterns. The system ingests organizational communication streams (e.g. emails, calendars, chat logs) to automatically generate activity diagrams mapping relationships between actors, tools, rules, and community structures for specific tasks. Within these activity frameworks, detailed persona profiles serve as persistent social memory, encoding individual social positioning, behavioral patterns, and contextual knowledge. Rather than relying on computationally expensive fine-tuning for constantly evolving organizational contexts, the system employs RAG to dynamically access relevant personas and activity models when processing organizational requests. We further propose to explore both centralized agents with global access to organizational knowledge and decentralized personal agents with selective information sharing across agents, to examine tradeoffs between social competence, privacy, and adaptability. This approach enables socially-informed reasoning through prompt-based instruction augmented with retrieved social context, with potential enhancement through reinforcement learning from AI feedback [2] to optimize social appropriateness over time.

To explore and validate this approach, we propose a three-stage experimental plan, comparing centralized and decentralized agent architectures across key metrics of social competence and coordination effectiveness. Stage 1 will develop controlled synthetic scenarios of liquidity crisis management in asset management firms, grounded in regulatory frameworks (e.g., SEC guidelines, Basel III [3, 18]) and industry case studies. This setting enables expert-annotated ground-truth evaluation of persona accuracy, activity model completeness, and task execution effectiveness, providing a foundation for measuring the utility of structured social representations in complex coordination tasks. Stage 2 will apply the framework to the Enron email corpus [10] to assess robustness in real-world organizational data, while acknowledging limitations of the dataset's temporal scope (2000-2002), potential selection biases, and privacy concerns. Evaluation metrics will include hierarchy detection accuracy, communication pattern recognition, social role identification, and representation coherence, supported by network analysis and expert qualitative review. Stage 3 will involve controlled deployment within a live asset management firm's risk department, with architectural choice (centralized or decentralized) determined in collaboration with organizational stakeholders based on privacy, data access, and coordination efficiency needs. Success metrics will combine quantitative measures—task completion rates, information retrieval accuracy, social appropriateness ratings—with qualitative assessments of perceived competence, trust, coordination fluency, and organizational integration. Finally, we will compare how centralized and decentralized frameworks shape the evolution and effectiveness of social representations, to inform practical deployment strategies under real-world privacy and security constraints.

2 RELATED WORK

While significant progress has been made in developing socially capable AI agents, much of this work remains focused on multi-agent coordination in controlled environments. Understanding how current approaches map to real-world organizational contexts reveals both promising techniques and key limitations.

Social Competence in AI Agents: Advances in large language models (LLMs) have enabled AI agents with emerging capabilities in reasoning, planning, and decision-making across diverse domains [4, 14, 19]. However, deploying these models in dynamic social contexts remains challenging due to shifting goals, ambiguous objectives, and evolving norms [5]. Effective social AI agents require coordination, negotiation, and adaptation while navigating cultural expectations and institutional structures [6].

Current Architectures and Capabilities: Advances in LLMs have enabled emerging capabilities in reasoning, planning, and social interaction across diverse domains [1, 4, 9, 14, 15, 19]. Hybrid systems combine LLMs with RL, strategic reasoning modules, or constraint-based optimization to model long-term behavior [11, 16, 20]. Agents demonstrate promising coordination and negotiation in structured environments, such as Diplomacy [9], Smallville [15], and simulated resource planning [16, 20]. However, these successes occur in domains with explicit rules and well-defined objectives. Organizational competence requires reasoning over implicit norms, tacit social cues, and dynamic power structures [5].

Evaluation Environments and Limitations: Most studies benchmark agents in controlled simulations—Smallville [15, 17], GOVSIM [16], economic games [20]—or in short-term, laboratory-style experiments [9, 12]. Few are evaluated within authentic organizational settings involving power structures, informal communication protocols, and long-term trust dynamics. Evaluation windows remain episodic, lacking sustained measures of trust formation, cross-context adaptation, or institutional integration.

Persistent Challenges for Organizational Integration: Despite progress in emergent social behavior, memory-based reasoning, and norm following [12, 15, 17], existing agents lack systematic frameworks for acquiring and reasoning about situated organizational knowledge: relationships, communication practices, project histories, influence networks. This gap motivates our approach to building persistent organizational memory and activity-based reasoning to support long-term integration of AI agents in evolving social contexts.

3 ORGANIZATIONAL SCENARIO: LIQUIDITY CRISIS IN ASSET MANAGEMENT

An ideal testbed for socially capable AI agents must expose the core challenges of organizational embeddedness: navigating conflicting stakeholder goals, negotiating hierarchies, adapting to informal norms, and responding to evolving institutional pressures. Organizational crises, where legal, market, and reputational risks converge, are particularly revealing. In such contexts, agents must coordinate with diverse actors under time pressure, interpret tacit cues, and mediate competing priorities while respecting roles and constraints. We model a liquidity crisis at ACME Asset Management, a fictional $12B institutional investment firm, designed to surface the

social reasoning required for effective organizational action under pressure.

A liquidity crisis occurs when client withdrawals exceed available cash, often triggered by negative news or market fears. Forced asset sales may follow, locking in losses and deepening mistrust. If confidence erodes, further withdrawals escalate the crisis. Beyond financial impact, such events risk regulatory violations, legal penalties, and reputational damage.

The crisis creates immediate conflict between departments. Risk Management urges rapid asset sales to meet regulatory liquidity thresholds. The Portfolio Manager resists, seeking to avoid performance losses. Debt financing is proposed but increases exposure if outflows persist.

Meanwhile, Sales need clear updates to stabilize client trust. But Financial Analysts, under internal pressure, must first run complex scenarios (asset sales, debt impacts), delaying actionable insights. Simultaneously, Board members and senior executives press for updates. Yet internal teams face reputational risks: no group wants to appear disorganized or responsible for delays. Each seeks to "own" positive outcomes, but premature or unaligned communication could harm credibility with clients, regulators, and the board.

Timing is critical. Analysts must prioritize: support Sales to reassure clients, or fulfill internal reporting? Leadership must sequence decisions and communications to maintain organizational coherence and public credibility.

This scenario highlights a core challenge for embedded AI agents: understanding how social structures, timing, and institutional logics shape effective action—not merely processing data, but reasoning about when, how, and through whom to act within complex organizational dynamics.

4 RESEARCH AGENDA: MODELING SOCIAL KNOWLEDGE FOR ORGANIZATIONALLY SITUATED AI AGENTS

To operationalize the proposed scenario, we will generate a synthetic organizational dataset based on structured Activity Theory diagrams and dynamic Persona profiles for key roles in the liquidity crisis: Risk Manager, Portfolio Manager, Financial Analysts, Sales/Advisors, Legal Counsel, Senior Executives, and Board members. The Activity Theory map will capture goals, roles, tools, norms, and community structures that govern decision-making in this context, grounded in regulatory frameworks (e.g., SEC guidelines, Basel III [3, 18]) and industry case studies.

Using this structure, we will create a corpus of fictional emails and chat logs simulating the unfolding crisis. Prompts will guide large language models to generate communications for each Persona, reflecting role-specific priorities, incentives, and institutional constraints. To explore agent variability and robustness, we will run multiple agent trials on the same scenario—varying generation parameters (e.g., temperature, sampling)—to produce diverse communication trajectories and decision paths. This will enable analysis of how social reasoning evolves across runs, and how Activity Theory modeling influences consistency and adaptability.

AI agents will then be tasked with reconstructing Personas and Activity Theory representations over time from this communication corpus, supporting their ability to reason about evolving social dynamics, institutional roles, and interaction norms. Agents will be evaluated on their capacity to guide appropriate communications and decisions in the simulated organizational environment.

To explore architectural trade-offs, we will compare two frameworks:

Centralized Agent Framework: A single agent with global access to communications and unified social memory, responsible for advising on actions across roles.

Decentralized Agent Framework: Separate agents assigned to individual Personas, each maintaining partial, role-specific social memory and coordinating through inter-agent messaging aligned with organizational norms.

Agents will be tested both with and without explicit Activity Theory-based modeling, to evaluate its contribution to reasoning accuracy, adaptability, and communication effectiveness. Key comparisons will focus on: (1) quality and coherence of reconstructed social knowledge; (2) ability to support role-appropriate actions under time pressure; (3) adaptability to evolving organizational dynamics; and (4) trade-offs in privacy, explainability, and coordination overhead. This agenda aims to advance understanding of how AI agents can model and act upon complex social knowledge in real organizational contexts.

REFERENCES

[1] 2024. AI can help humans find common ground in democratic deliberation. 386 (Oct. 2024), eadq2852. https://doi.org/10.1126/science.adq2852

[2] Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, Carol Chen, Catherine Olsson, Christopher Olah, Danny Hernandez, Dawn Drain, Deep Ganguli, Dustin Li, Eli Tran-Johnson, Ethan Perez, Jamie Kerr, Jared Mueller, Jeffrey Ladish, Joshua Landau, Kamal Ndousse, Kamile Lukosuite, Liane Lovitt, Michael Sellitto, Nelson Elhage, Nicholas Schiefer, Noemi Mercado, Nova DasSarma, Robert Lasenby, Robin Larson, Sam Ringer, Scott Johnston, Shauna Kravec, Sheer El Showk, Stanislav Fort, Tamera Lanham, Timothy Telleen-Lawton, Tom Conerly, Tom Henighan, Tristan Hume, Samuel R. Bowman, Zac Hatfield-Dodds, Ben Mann, Dario Amodei, Nicholas Joseph, Sam McCandlish, Tom Brown, and Jared Kaplan. 2022. Constitutional AI: Harmlessness from AI Feedback. arXiv:2212.08073 [cs.CL] https://arxiv.org/abs/2212.08073

[3] Basel Committee on Banking Supervision. 2010. Basel III: A global regulatory framework for more resilient banks and banking systems. https://www.bis.org/publ/bcbs189.pdf.

[4] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. arXiv:2005.14165 [cs.CL] https://arxiv.org/abs/2005.14165

[5] Cédric Colas, Tristan Karch, Clément Moulin-Frier, and Pierre-Yves Oudeyer. 2022. Language and culture internalization for human-like autotelic AI. *Nature Machine Intelligence* 4, 12 (Dec. 2022), 1068–1076. https://doi.org/10.1038/s42256-022-00591-4

[6] Allan Dafoe, Edward Hughes, Yoram Bachrach, Tantum Collins, Kevin R. McKee, Joel Z. Leibo, Kate Larson, and Thore Graepel. 2020. Open Problems in Cooperative AI. arXiv:2012.08630 (Dec. 2020). https://doi.org/10.48550/arXiv.2012.08630 arXiv:2012.08630 [cs].

[7] Sebastian Döweling, Benedikt Schmidt, and Andreas Göb. 2012. A model for the design of interactive systems based on activity theory. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (Seattle, Washington, USA) *(CSCW '12)*. Association for Computing Machinery, New York, NY, USA, 539–548. https://doi.org/10.1145/2145204.2145287

[8] Yrjö Engeström. 2000. Activity theory as a framework for analyzing and redesigning work. *Ergonomics* 43, 7 (2000), 960–974. https://doi.org/10.1080/001401300409143

[9] Meta Fundamental AI Research Diplomacy Team (FAIR)†, Anton Bakhtin, Noam Brown, Emily Dinan, Gabriele Farina, Colin Flaherty, Daniel Fried, Andrew Goff, Jonathan Gray, Hengyuan Hu, Athul Paul Jacob, Mojtaba Komeili, Karthik Konath, Minae Kwon, Adam Lerer, Mike Lewis, Alexander H. Miller, Sasha Mitts, Adithya

Renduchintala, Stephen Roller, Dirk Rowe, Weiyan Shi, Joe Spisak, Alexander Wei, David Wu, Hugh Zhang, and Markus Zijlstra. 2022. Human-level play in the game of Diplomacy by combining language models with strategic reasoning. *Science* 378, 6624 (Dec. 2022), 1067–1074. https://doi.org/10.1126/science.ade9097

[10] Bryan Klimt and Yiming Yang. 2004. The enron corpus: a new dataset for email classification research. In *Proceedings of the 15th European Conference on Machine Learning* (Pisa, Italy) *(ECML'04)*. Springer-Verlag, Berlin, Heidelberg, 217–226. https://doi.org/10.1007/978-3-540-30115-8_22

[11] Connor Lawless, Jakob Schoeffer, Lindy Le, Kael Rowan, Shilad Sen, Cristina St. Hill, Jina Suh, and Bahareh Sarrafzadeh. 2024. "I Want It That Way": Enabling Interactive Decision Support Using Large Language Models and Constraint Programming. arXiv:2312.06908 [cs.HC] https://arxiv.org/abs/2312.06908

[12] Soohwan Lee, Seoyeong Hwang, Dajung Kim, and Kyungho Lee. 2025. Conversational Agents as Catalysts for Critical Thinking: Challenging Social Influence in Group Decision-making. arXiv:2503.14263 (March 2025). https://doi.org/10.48550/arXiv.2503.14263 arXiv:2503.14263 [cs].

[13] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2021. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. arXiv:2005.11401 [cs.CL] https://arxiv.org/abs/2005.11401

[14] Jiachang Liu, Dinghan Shen, Yizhe Zhang, Bill Dolan, Lawrence Carin, and Weizhu Chen. 2021. What Makes Good In-Context Examples for GPT-3? arXiv:2101.06804 [cs.CL] https://arxiv.org/abs/2101.06804

[15] Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. arXiv:2304.03442 (Aug. 2023). https://doi.org/10.48550/arXiv.2304.03442 arXiv:2304.03442 [cs].

[16] Giorgio Piatti, Zhijing Jin, Max Kleiman-Weiner, Bernhard Schölkopf, Mrinmaya Sachan, and Rada Mihalcea. 2024. Cooperate or Collapse: Emergence of Sustainable Cooperation in a Society of LLM Agents. arXiv:2404.16698 (Dec. 2024). https://doi.org/10.48550/arXiv.2404.16698 arXiv:2404.16698 [cs].

[17] Siyue Ren, Zhiyao Cui, Ruiqi Song, Zhen Wang, and Shuyue Hu. 2024. Emergence of Social Norms in Generative Agent Societies: Principles and Architecture. In *Proceedings of the Thirty-ThirdInternational Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, Jeju, South Korea, 7895–7903. https://doi.org/10.24963/ijcai.2024/874

[18] U.S. Securities and Exchange Commission. 2016. Investment Company Liquidity Risk Management Programs, Final Rule. https://www.sec.gov/rules/final/2016/33-10233.pdf. Release No. 33-10233; IC-32315; File No. S7-16-15.

[19] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed H. Chi, Quoc Le, and Denny Zhou. 2022. Chain of Thought Prompting Elicits Reasoning in Large Language Models. *CoRR* abs/2201.11903 (2022). arXiv:2201.11903 https://arxiv.org/abs/2201.11903

[20] Stephan Zheng, Alexander Trott, Sunil Srinivasa, David C. Parkes, and Richard Socher. 2021. The AI Economist: Optimal Economic Policy Design via Two-level Deep Reinforcement Learning. arXiv:2108.02755 (Aug. 2021). https://doi.org/10.48550/arXiv.2108.02755 arXiv:2108.02755 [cs].