Multi-Agent Systems for Information Systems Research: A Framework for Collaborative AI-Augmented Inquiry

Research in Progress

Benjamin M. Ampel
Georgia State University
bampel@gsu.edu

Steven Ullman
University of Texas, San Antonio
Steven.ullman@utsa.edu

Abstract

Contemporary Information Systems research demands unprecedented interdisciplinary expertise and systematic literature integration. These requirements strain individual researchers' cognitive resources. This work-in-progress presents the Multi-Agent Systems for Information Systems Research (MAS4ISR) framework, which addresses these challenges through theoretically grounded collaborative AI agents. The framework operationalizes transactive memory theory and coordination theory to enable autonomous Artificial Intelligence (AI) agents to conduct literature synthesis, analysis, and synthesis tasks. Unlike existing multi-agent frameworks, MAS4ISR implements a computational transactive memory system and evidence-based consensus mechanisms to enable the coordination of specialized AI agents. The framework can potentially offer theoretical contributions to IS research and practical value for AI-augmented scholarly inquiry.

Keywords: Multi-Agent Systems, Information Systems Research, Artificial Intelligence, Transactive Memory Theory, Coordination Theory, Research Automation

Introduction

The Information Systems (IS) discipline requires interdisciplinary expertise, systematic literature integration, and sophisticated methodological rigor that often exceed individual researchers' cognitive and temporal capacities. This challenge is particularly acute in emergent domains such as artificial intelligence (AI), cybersecurity, and blockchain, where rapid technological evolution outpaces traditional academic knowledge production cycles. Autonomous AI agents offer a potential solution to these resource constraints Lambiase et al. 2024). Unlike conventional decision support systems, AI agents can execute complex tasks autonomously (Lu et al. 2025). However, adapting AI agents to IS research presents non-trivial challenges. While individual AI agents demonstrate strong domain-specific capabilities, they lack the sophisticated coordination mechanisms characterizing effective human research teams. To address these limitations, we propose the Multi-Agent Systems for Information Systems Research (MAS4ISR) framework. MAS4ISR systematically operationalizes organizational theory in computational systems through four key innovations:

1. A computational transactive memory system enabling shared knowledge structures.
2. Evidence-based consensus mechanisms enforcing academic standards.
3. Systematic integration of coordination and transactive memory theories.
4. A research-specific quality evaluation framework combining automated content analysis with subjective assessment.

MAS4ISR can be conceptualized as an agentic service system in which autonomous AI agents act as service actors that co-create value within a knowledge-production ecosystem (Skodawessely et al. 2024; Zhao & Zaki 2024). The framework's transactive memory and coordination mechanisms parallel those found in hybrid human-AI service architectures, where effective orchestration depends on shared expertise structures and adaptive coordination (Skodawessely et al. 2024; Zhao & Zaki 2024). By operationalizing these mechanisms, MAS4ISR advances the notion of AI-enabled service orchestration and facilitates knowledge-as-a-service interactions that enhance collective research capability among human and machine agents. This work-in-progress can also potentially contribute to IS theory and practice by demonstrating how established organizational theories can inform autonomous AI agent design while providing a systematic approach to AI-augmented research that maintains theoretical rigor and methodological sophistication (Gregor and Hevner 2013; Hevner et al. 2004; Larsen et al. Forthcoming).

Theoretical Foundations and Related Work

AI Agents in Information Systems Research

Early IS contributions focused on expert and knowledge-based decision support systems that augmented human decision-making under human supervision (Xu et al. 2014). While successful in narrow domains, these systems lacked the flexibility and adaptability required for complex organizational tasks. The IS discipline has therefore started to transition from passive decision support systems to active collaborative intelligence (Stelmaszak et al. 2025). Research has demonstrated that autonomous AI agents can exhibit goal-directed behavior, contextual reasoning, and adaptive capabilities with-out explicit human guidance (Keogh and Sonenberg 2020). This evolution aligns with IS literature emphasizing how AI systems are best suited for enhancing organizational capabilities (Lin and Maruping 2025). However, effective multi-agent coordination requires sophisticated mechanisms that manage interdependencies among autonomous agents while adapting to changing organizational needs (Borghoff et al. 2025).

Theoretical Foundations: Transactive Memory and Coordination

Organizational theory provides frameworks for understanding collaborative knowledge work. Transactive memory theory explains how teams develop shared knowledge structures that enable specialization and coordination (Weigel et al. 2012). Transactive memory theory has demonstrated that when teams understand each other's capabilities and expertise, members can distribute cognitive load and access specialized expertise when needed (He et al. 2022). IS research has extensively applied transactive memory theory to distributed teams and knowledge management systems (Nevo and Wand 2005; Oshri et al. 2008; Zhang et al. 2025). Coordination theory elucidates how autonomous AI agents can potentially manage interdependencies and achieve collective goals (Crowston et al. 2015). This includes identifying specific mechanisms (standardization, planning, and mutual adjustment) that enable effective collaboration under different task conditions (Crowston et al. 2015). These theoretical foundations provide insights for designing multi-agent systems that address complex collaborative challenges in organizational settings.

The MAS4ISR Framework

Framework Overview

This research proposes the MAS4ISR framework, which represents a modular multi-agent architecture designed to automate and support end-to-end IS research ideation, literature synthesis, and proposal generation. Our framework will address identified research gaps through systematic integration of theoretical foundations with practical AI technologies (Samtani et al. 2023). Extant multi-agent frameworks lack persistent memory and theoretical grounding (Hong et al. 2024; Qian et al. 2024; Wu et al. 2024). Our proposed system will implement three specialized agent types. First, a research agent for literature search and keyword generation. Second, an analysis agent for data analysis and insight extraction. Finally, a synthesis agent for proposal writing and quality assessment. Each agent performs iterative reasoning-action cycles, reasoning about the current state, planning actions, executing those actions, and observing results before proceeding (Yao et al. 2023). Agents maintain integration with the Transactive Memory System, updating their expertise levels based on task performance and accessing shared knowledge for improved decision-making. Figure 1 illustrates the overall architecture of the MAS4ISR framework, showing how specialized agents interact through the transactive memory system to achieve coordinated research outcomes.

Transactive Memory System

The MAS4ISR framework operationalizes transactive memory theory to enable efficient knowledge distribution and retrieval across specialized agents. The MAS4ISR framework is defined as:

$$\mathrm{TMS} = (\mathrm{M}, \mathrm{D}, \mathrm{R})$$

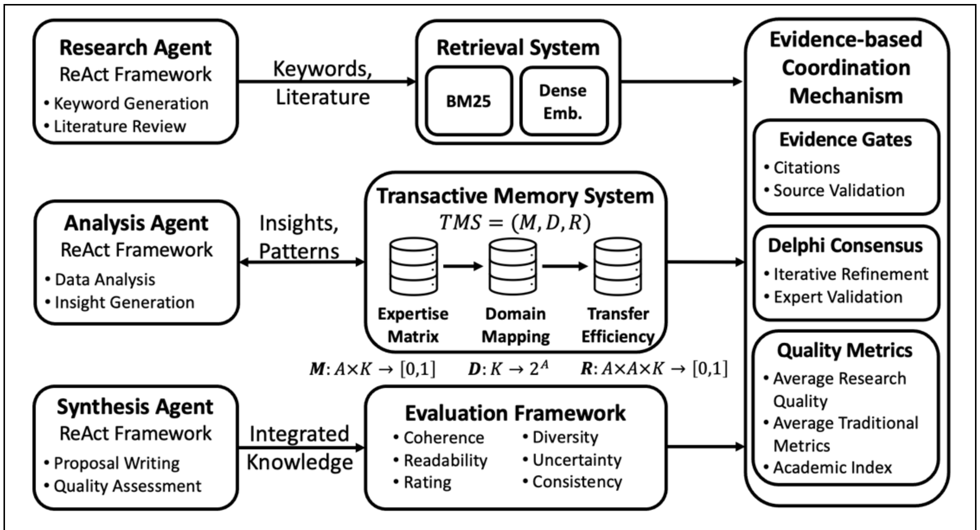

Figure 1. Proposed MAS4ISR Framework

where A is the set of agents and K is the set of knowledge domains. For each agent-knowledge pair $(a_i, k_j) \in A \times K$, the function $M(a_i, k_j) \in [0,1]$ represents the expertise level of agent a_i in knowledge domain k_j. The directory function $D: K \to 2^A$ identifies which agents possess knowledge in each domain, while $R: A \times A \times K \to [0,1]$ determines how efficiently knowledge can be transferred between agent pairs for a given domain. The framework implements expertise learning through performance-based updates:

$$M(a_i, k_j)^{t+1} = \alpha \cdot M(a_i, k_j)^t + (1 - \alpha) \cdot \mathrm{performance}(a_i, k_j)$$

where $\alpha \in [0,1]$ is the learning rate and $\mathrm{performance}(a_i, k_j) \in [0,1]$ measures the quality of agent a_i's contributions in domain k_j, computed as the normalized average of relevance scores for documents retrieved by the agent in that domain. Knowledge items are represented as paper objects containing title, abstract, authors, and venue information. The system employs a hybrid retrieval approach combining BM25 keyword-based ranking with dense embedding semantic similarity for effective knowledge access across diverse research contexts. This approach builds upon recent advances in memory-augmented systems while incorporating organizational theory foundations absent in existing frameworks (Packer et al. 2024).

Coordination System

The MAS4ISR framework implements coordination through a sequential workflow-based architecture combined with consensus-based decision making. This approach addresses the fundamental challenge of coordinating multiple specialized agents to achieve coherent research outcomes while maintaining individual agent autonomy and expertise. Our framework executes a structured coordination protocol with eight steps: (1) keyword generation, (2) literature search, (3) evidence hunting, (4) methodology planning, (5) data collection planning, (6) proposal synthesis, (7) quality improvement, and (8) consensus building. Each phase produces structured outputs that serve as inputs for subsequent phases. This sequential coordination ensures that each agent has access to the complete context from previous phases while maintaining clear responsibility boundaries.

For decision-making and quality assurance, the system implements Delphi consensus mechanisms to achieve expert validation through iterative feedback. This consensus process operates by conducting multi-round expert assessments where agents independently evaluate outputs, analyze areas of agreement and disagreement, and iteratively refine their judgments until convergence on a consensus threshold. This approach differs from adversarial debate frameworks by emphasizing convergence rather than divergence through structured consensus building (Liang et al. 2024).

Preliminary Evaluation and Results

Evaluation Methodology

Our evaluation framework assesses AI approaches for IS research tasks along three complementary dimensions (Ampel et al. 2025; Kara et al. 2025). First, we provide a traditional metrics average that integrates measures of discourse coherence, diversity, embedding-based consistency, and model-based holistic judgment. DiscoScore employs BERT-based evaluation for text consistency. Diversity measures inter-text semantic variation with $\text{Diversity} = 1 - \frac{1}{n(n-1)}\sum_{i \neq j} \text{cosine\_similarity}(e_i, e_j)$. Coherence measures intra-text semantic consistency with $\text{Coherence} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i(m_i-1)}\sum_{j \neq k} \text{cosine\_similarity}(s_{ij}, s_{ik})$. The LLM judge score prompts GPT-4o to evaluate the overall quality of outputs across multiple dimensions, including relevance, rigor, and feasibility with $LJS = \frac{1}{n}\sum_{i=1}^{n} \text{LLM\_Score}_i$. The overall traditional metrics average is calculated as:

$$TMA = \frac{DS + \frac{LJS}{10} + C + D}{4}$$

Second, the Academic Writing Index synthesizes readability indices to assess text complexity: $AWI = 0.4 \times GF + 0.3 \times SMOG + 0.2 \times ARI + 0.1 \times CL$, where scores exceeding 70 indicate superior academic writing quality. However, traditional linguistic metrics inadequately capture domain-specific research quality dimensions. To address this limitation, we developed the Overall Research Quality (ORQ) score as a measure integrating six critical research dimensions: literature grounding, methodological rigor, technical depth, academic structure, novelty, and reproducibility. Each dimension employs a hybrid approach combining automated pattern matching with GPT-4o-based assessment. Literature grounding and methodological rigor use LLM-as-a-judge evaluation for citation quality and research design assessment. Technical depth and academic structure employ automated pattern matching for technical indicators and section completeness. Novelty implements LLM-as-a-judge evaluation for research contribution assessment. Reproducibility uses automated pattern matching to count implementation-related keywords (code, dataset, algorithm, parameter, etc.). The ORQ score is computed as the research metrics average:

$$ORQ = \frac{1}{6}\sum_{i=1}^{6} S_i\, w_i$$

where S_i represents individual dimension scores and w_i their corresponding weights. This multi-dimensional evaluation framework provides a comprehensive assessment across research quality, linguistic measures, and academic writing standards, enabling systematic comparison of AI approaches for IS research tasks.

Preliminary Results

To evaluate our MAS4ISR, we evaluated it against zero-shot, few-shot, Chain-of-Thought reasoning, and self-consistency prompting on GPT-4o. These benchmarks are simple baselines widely used across the literature for evaluating AI system performance in research and knowledge work tasks (Ampel et al. 2025). Zero-Shot and few-Shot approaches represent the most common deployment patterns for LLMs in academic contexts. Chain-of-Thought and self-consistency methods represent enhanced reasoning techniques that have demonstrated improved performance in complex reasoning tasks (Liu et al. 2023). Table 1 presents baseline comparison results across multiple IS research domains.

Model	ORQ	Traditional	AWI
Zero-Shot GPT-4o	0.340***	0.443	19.527
Few-Shot GPT-4o	0.319***	0.420	19.386

Chain-of-Thought GPT-4o	0.353^{***}	0.433	18.116
Self-Consistency GPT-4o	0.333^{***}	0.426	18.684
MAS4ISR	0.494	0.380	16.149

Table 1. A Very Nice Table

MAS4ISR achieved the highest research quality (0.494) compared to all baseline models, with the closest competitor being Chain-of-Thought GPT-4o (0.353). This substantial advantage in research-specific metrics suggests that multi-agent coordination and transactive memory mechanisms provide meaningful improvements in literature grounding, methodological rigor, and technical depth assessment. However, an interesting pattern emerges in traditional linguistic metrics, where Zero-Shot GPT-4o achieves the highest traditional metrics average (0.443), followed by Chain-of-Thought GPT-4o (0.433), while MAS4ISR shows lower performance (0.380). This counterintuitive result may reflect that zero-shot prompting preserves the model's inherent linguistic capabilities without the potential interference from few-shot examples or reasoning chains that might introduce inconsistencies. Zero-Shot GPT-4o also achieves the highest academic writing index (19.527), suggesting that unconstrained generation may better preserve the model's training on academic text patterns. These results demonstrate that our MAS4ISR framework excels in research-specific quality dimensions but traditional prompting strategies may still hold advantages in linguistic coherence and academic writing style.

Discussion and Implications

This work-in-progress will contribute to IS theory by demonstrating how established organizational theories can inform the design of AI agents (Borghoff et al. 2025). The MAS4ISR framework can potentially address the gap between organizational science and AI system design by demonstrating how insights about human collaboration can enhance artificial agent coordination. This contributes to the growing body of IS literature on AI-augmented knowledge work (Ampel et al. 2025; Samtani et al. 2023) and provides a model for theoretically grounded AI system development (Rai 2017). From a practical perspective, the MAS4ISR framework will offer a systematic approach to AI-augmented research that can enhance productivity, quality, and innovation in IS investigations. The framework will address the growing complexity of IS research while maintaining the theoretical rigor and methodological sophistication that characterize high-quality scholarship. Following the completion of MAS4ISR, we will conduct a user study with IS researchers to evaluate the framework's practical utility and identify areas for improvement. This user study will provide valuable feedback on the framework's usability, effectiveness, and potential areas for improvement.

Captions should be using the "Caption" style – configured as Georgia 10-point bold. They should be numbered (e.g., "Table 1" or "Figure 2"), centered and placed beneath the figure or table. Please note that the words "Figure" and "Table" should be spelled out (e.g., "Figure" rather than "Fig.") wherever they occur. The proceedings will be made available online, thus color figures are possible.

Conclusion

This work-in-progress presents the MAS4ISR framework, a theoretically grounded approach to multi-agent AI collaboration for IS research. Our framework addresses limitations in extant multi-agent systems through the integration of transactive memory theory and coordination theory, providing practical utility for AI-augmented scholarly inquiry. Our proposed MAS4ISR framework provides a computational instantiation of transactive memory theory that enables persistent knowledge structures and dynamic expertise allocation, an evidence-based consensus mechanisms that enforce academic standards and reduce hallucinations, systematic integration of organizational theories with multi-agent AI design, and a research-specific evaluation framework that measures academic quality across multiple dimensions. Current limitations include the need for more extensive empirical validation across diverse research domains and longer-term studies of system performance. Future work will extend the framework to broader domains of AI-augmented knowledge work, incorporate additional organizational theories, and conduct comprehensive user studies with IS researchers to validate practical utility and identify areas for improvement. Future work will include a user study to evaluate the framework's practical utility and identify areas for improvement.

References

Ampel, B., Yang, C.-H., Hu, J., and Chen, H. (2025). "Large language models for conducting advanced text analytics information systems research," ACM Transactions on Management Information Systems (16:1), pp. 1–27.

Borghoff, U. M., Bottoni, P., and Pareschi, R. (2025). "An organizational theory for multi-agent interactions integrating human agents, LLMs, and specialized AI," Discover Computing (28:1), pp. 1–35.

Crowston, K., Rubleske, J., and Howison, J. (2015). "Coordination theory: A ten-year retrospective," in Human-computer interaction and management information systems: Foundations, Routledge, pp. 134–152.

Gregor, S. and Hevner, A. R. (2013). "Positioning and presenting design science research for maximum impact," MIS Quarterly (), pp. 337–355.

He, W., Po-An Hsieh, J., Schroeder, A., and Fang, Y. (2022). "Attaining individual creativity and performance in multidisciplinary and geographically distributed IT project teams: the role of Transactive memory systems," MIS Quarterly (46:2), pp. 1035–1072.

Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). "Design science in information systems research," MIS Quarterly (), pp. 75–105.

Hong, S., Zhuge, M., Chen, J., Zheng, X., Cheng, Y., Zhang, C., Wang, J., Wang, Z., Yau, S. K. S., Lin, Z., et al. (2024). "MetaGPT: Meta Programming for a Multi-Agent Collaborative Framework," in 12th International Conference on Learning Representations, ICLR 2024,

Kara, M., Ozduran, E., Kara, M. M., Özbek, İ. C., and Hancı, V. (2025). "Evaluating the readability, quality, and reliability of responses generated by ChatGPT, Gemini, and Perplexity on the most commonly asked questions about Ankylosing spondylitis," PLoS One (20:6), e0326351.

Keogh, K. and Sonenberg, L. (2020). "Designing multi-agent system organisations for flexible runtime behaviour," Applied Sciences (10:15), p. 5335.

Lambiase, S., Catolino, G., Palomba, F., and Ferrucci, F. (2024). "Motivations, Challenges, Best Practices, and Benefits for Bots and Conversational Agents in Software Engineering: A Multivocal Literature Review," ACM Computing Surveys (57:4), pp. 1–37.

Larsen KR, Lukyanenko R, Mueller RM, et al (2025) Validity in Design Science. MIS Quarterly, Forthcoming.

Liang, T., He, Z., Jiao, W., Wang, X., Wang, Y., Wang, R., Yang, Y., Shi, S., and Tu, Z. (2024). Encouraging Divergent Thinking in Large Language Models through Multi-Agent Debate," in Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, pp. 17889–17904.

Lin, Y.-K. and Maruping, L. M. (2025). "Organizing for AI Innovation: Insights From an Empirical Exploration of US Patents," MIS Quarterly (49:3), pp. 1095–1122.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., and Neubig, G. (2023). "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing," ACM Computing Surveys (55:9), pp. 1–35.

Lu, Y., Yang, S., Qian, C., Chen, G., Luo, Q., Wu, Y., Wang, H., Cong, X., Zhang, Z., Lin, Y., Liu, W., Wang, Y., Liu, Z., Liu, F., and Sun, M. (2025). "Proactive Agent: Shifting LLM Agents from Reactive Responses to Active Assistance," in International Conference on Learning Representations (ICLR) 2025 Workshop / Poster Track, Poster.

Nevo, D. and Wand, Y. (2005). "Organizational memory information systems: a transactive memory approach," Decision Support Systems (39:4), pp. 549–562.

Oshri, I., Van Fenema, P., and Kotlarsky, J. (2008). "Knowledge transfer in globally distributed teams: the role of transactive memory," Information Systems Journal (18:6), pp. 593–616.

Packer, C., Martinez, V., Yang, R., Frazier, J., Weinstein, J., MacGlashan, J., and Narasimhan, K. (2024). "MemGPT: Towards LLMs as Operating Systems," arXiv preprint arXiv:2310.08560 ().

Qian, C., Liu, W., Liu, H., Chen, N., Dang, Y., Li, J., Yang, C., Chen, W., Su, Y., Cong, X., et al. (2024). "ChatDev: Communicative Agents for Software Development," in Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 15174–15186.

Rai, A. (2017). "Editor's Comments: Diversity of Design Science Research," Management Information Systems Quarterly (41:1), pp. iii–xviii.

Samtani, S., Zhu, H., Padmanabhan, B., Chai, Y., Chen, H., and Nunamaker Jr, J. F. (2023). "Deep learning for information systems research," Journal of Management Information Systems (40:1), pp. 271–301.

Skodawessely, T., Jung, H., & Zinke-Wehlmann, C. (2025). Service-embedded interaction for LLM assistance. In Proceedings of the Pre-ICIS SIG Services Workshop on Synergizing Service Ecosystems with AI, Bangkok, Thailand, 2024.

Stelmaszak, M., Möhlmann, M., and Sørensen, C. (2025). “When algorithms delegate to humans: Exploring human-algorithm interaction at Uber,” MIS Quarterly (49:1), pp. 305–330.

Weigel FK, Ezell JD, Hazen BT (2012). “Transactive memory theory: A review of the literature and suggestions for future MIS research,” Proceedings of the southern association for information systems conference, pp. 1–6.

Wu, Q., Bansal, G., Zhang, J., Wu, Y., Li, B., Zhu, E., Jiang, L., Zhang, X., Zhang, S., Liu, J., et al. (2024). “Autogen: Enabling next-gen LLM applications via multi-agent conversations,” in First Conference on Language Modeling,

Xu, J., Benbasat, I., and Cenfetelli, R. T. (2014). “The nature and consequences of trade-off transparency in the context of recommendation agents,” MIS quarterly (38:2), pp. 379–406.

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., and Cao, Y. (2023). “React: Synergizing reasoning and acting in language models,” in International Conference on Learning Representations (ICLR).

Zhang, X., Fang, Y., Zhou, J., and Lim, K. H. (2025). “How collaboration technology use affects IT project team creativity: Integrating team knowledge and creative synthesis perspectives,” MIS Quarterly (49:2), pp. 611–642.

Zhao, H., & Zaki, M. (2024). Optimizing service experience design with generative AI. In Proceedings of the Pre-ICIS SIG Services Workshop on Synergizing Service Ecosystems with AI, Bangkok, Thailand, 2024.