Vol-3 Issue-2 2025
Scientific Research Journal of Science, Engineering and Technology
ISSN: 2584-0584, Peer Reviewed Journal

Multi-Agent Systems for Action Item Extraction from Meeting Transcripts: A Comprehensive Review

Haruto Tanaka, Department of Artificial Intelligence and Data Science, Student, Kansai Institute of Advanced Science and Technology (KIAST), Japan

Prof. Grant Thompson, Department of Artificial Intelligence and Data Science, Professor, Kansai Institute of Advanced Science and Technology (KIAST), Japan

Abstract

The rapid growth of digital collaboration platforms has led to an unprecedented increase in recorded meetings and conversational data. These interactions are commonly preserved as textual transcripts through automated transcription technologies. While transcripts provide a complete record of discussions, their unstructured and verbose nature limits their direct usefulness for organizational follow-up and decision-making. Among the most valuable outcomes of meetings are action items, which capture tasks, responsibilities, and commitments that must be executed after the discussion ends. This review paper examines the role of multi-agent artificial intelligence systems in extracting action items from meeting transcripts. By synthesizing recent advances in multi-agent frameworks, large language models, conversational analysis, and meeting intelligence, the paper highlights how agent-based decomposition improves accuracy, interpretability, and scalability in action item extraction. The review also discusses architectural patterns, coordination strategies, and application contexts, offering a structured understanding of how multi-agent approaches address the limitations of traditional single-model pipelines.

Keywords

Multi-agent systems, meeting transcripts, action item extraction, large language models, document intelligence, conversational AI

1. Introduction

The digital transformation of organizations has fundamentally changed how collaboration and decision-making occur. Meetings conducted through video conferencing, teleconferencing, and collaborative platforms are routinely recorded and transcribed, generating large volumes of conversational text. Although automated transcription has become increasingly accurate, transcripts remain difficult to operationalize because they mix narrative discussion, opinions, explanations, and decisions within a single unstructured document. Researchers in conversation analysis and automated transcription have long highlighted the gap between raw transcripts and actionable organizational knowledge, emphasizing the need for intelligent post-processing mechanisms to support reflection, accountability, and follow-up activities (Oliver et al., 2005; Moore, 2015).

Action items represent a particularly critical form of extracted knowledge, as they encode commitments that directly influence project execution and organizational performance. However, identifying action

items is non-trivial because commitments are often implicit, distributed across multiple turns, or expressed using informal language. Recent advances in artificial intelligence adoption across professional and service-oriented industries demonstrate a growing demand for systems that can convert conversational data into structured outputs that support workflow automation and decision-making (Yang et al., 2024).

In parallel, multi-agent artificial intelligence has emerged as a powerful paradigm for addressing complex language understanding tasks. Instead of relying on a single monolithic model, multi-agent systems decompose problems into smaller, specialized subtasks handled by cooperating agents. Knowledge-driven and large language model–based agent frameworks have shown promising results in requirements engineering, qualitative analysis, cybersecurity problem solving, and conversational data synthesis (Jin et al., 2025; Huang et al., 2025; Xu et al., 2025). These developments motivate a focused review of how multi-agent approaches can be applied to meeting transcripts with the explicit goal of extracting action items. This paper reviews recent literature and synthesizes design principles for multi-agent action item extraction systems.

2. Background: Meeting Transcripts and Action Items

Meeting transcripts are a textual representation of spoken interaction, typically produced by automatic speech recognition systems. While they preserve conversational content, transcripts lack the structural markers needed for immediate interpretation, such as explicit task boundaries or decision labels. Research in transcription and qualitative analysis has shown that transcripts often require significant human interpretation to identify meaning, intent, and responsibility, especially in collaborative or high-stakes environments (Oliver et al., 2005; Moore, 2015).

Action items differ from general summaries in that they are inherently task-oriented. They describe activities that must be performed, often include an implicit or explicit owner, and may reference deadlines or conditions. Turn-taking dynamics, interruptions, and overlapping speech further complicate the extraction of such items, as commitments may be refined incrementally across multiple conversational turns (Patamia et al., 2025). Traditional rule-based or keyword-driven systems struggle to capture these nuances, leading to either missed actions or false positives.

Recent work on meeting intelligence and conversational visualization systems highlights the importance of accurately capturing actionable outcomes to enhance team communication and performance, particularly in complex or high-stress contexts (Oppermann et al., 2025). These insights underline the need for intelligent systems that can move beyond surface-level text processing toward deeper semantic and pragmatic understanding.

3. Emergence of Multi-Agent Approaches in Language Processing

Multi-agent systems have gained traction as a means of managing complexity in artificial intelligence applications. In this paradigm, multiple autonomous or semi-autonomous agents collaborate, each responsible for a specific aspect of a task. Knowledge-driven multi-agent frameworks have demonstrated effectiveness in requirements development by separating domain understanding, constraint reasoning, and validation into distinct agents that interact through shared representations (Jin et al., 2025).

With the integration of large language models, multi-agent architectures have expanded their applicability to open-ended reasoning and qualitative analysis tasks. For example, agent-based frameworks for

automated qualitative analysis divide responsibilities such as coding, theme identification, and validation among cooperating agents, improving consistency and transparency compared to single-model approaches (Xu et al., 2025). Similar principles have been applied in cybersecurity competitions and conversational data generation, where agents adopt complementary roles to solve complex problems or simulate realistic dialogue (Huang et al., 2025; Kirstein et al., 2025).

These studies collectively suggest that multi-agent systems are well suited for transcript analysis, where understanding intent, context, and commitment requires layered reasoning. By assigning distinct cognitive roles to agents, systems can better manage ambiguity and reduce the risk of hallucinated or unsupported outputs.

4. Multi-Agent Architectures for Action Item Extraction

Multi-agent architectures for action item extraction typically decompose the task into sequential or collaborative stages. An initial comprehension agent focuses on understanding the overall meeting context, participant roles, and topic flow. This agent establishes a shared semantic foundation that downstream agents rely upon, addressing challenges related to reference resolution and contextual continuity observed in conversational analysis research (Moore, 2015).

Subsequent agents are often responsible for intent and commitment detection. These agents analyze linguistic cues such as modality, future tense, imperatives, and agreement markers to identify utterances that potentially signal an action. Insights from turn-taking and conversational dynamics research inform the design of these agents, enabling them to track how commitments evolve across speaker turns (Patamia et al., 2025).

Action candidate extraction agents then transform detected intents into structured representations, capturing task descriptions, responsible entities, and temporal constraints when available. Validation agents play a critical role by filtering out speculative statements or suggestions that lack commitment, a challenge frequently noted in qualitative transcript analysis. Finally, consolidation agents merge duplicates and normalize language, producing a concise and actionable output suitable for integration with task management systems. Similar agent coordination patterns have been successfully applied in clinical documentation synthesis and privacy-preserving conversational generation, demonstrating their robustness across domains (van Velzen et al., 2025).

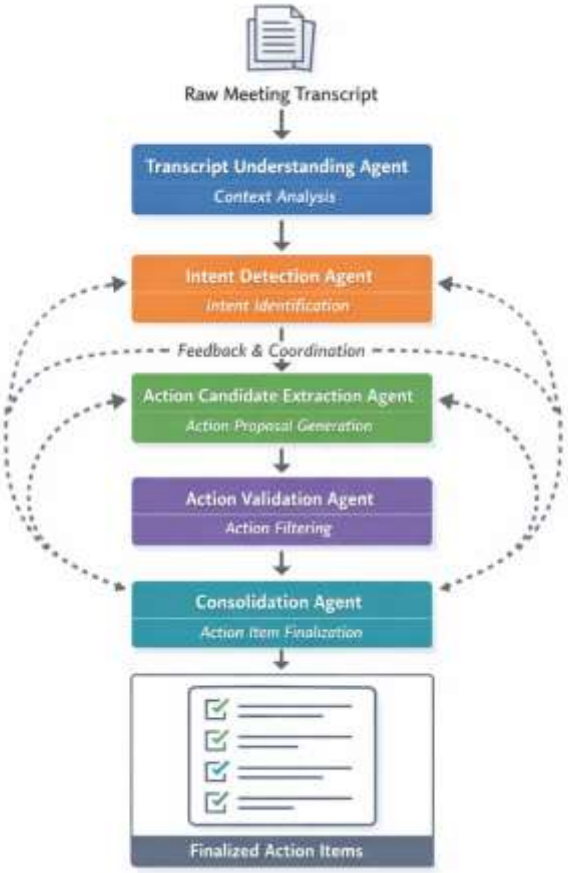

Figure 1: Multi-Agent Architecture for Action Item Extraction from Meeting Transcripts

The figure presents a layered, pipeline-style architecture that highlights how multi-agent collaboration enables accurate action item extraction from meeting transcripts. The process begins with the raw meeting transcript, which may contain speaker turns, disfluencies, and contextual dependencies. This transcript is first processed by a Transcript Understanding Agent, responsible for building a global semantic representation of the conversation, including topic flow and participant roles.

Next, an **Intent Detection Agent** analyzes utterances to identify expressions of commitment, obligation, or future-oriented actions using linguistic and pragmatic cues. Identified intents are passed to an **Action Candidate Extraction Agent**, which formulates preliminary action items in structured form, prioritizing recall over precision. These candidates are then evaluated by an **Action Validation Agent**, which filters out speculative statements, suggestions, or non-committal discussion to ensure that only executable actions are retained.

Finally, a **Consolidation Agent** merges redundant or overlapping actions, normalizes language, and produces a concise list of finalized action items. Throughout the architecture, **inter-agent communication channels** and **feedback loops** enable iterative refinement, allowing downstream agents to request clarification or reassessment from upstream agents. This coordinated reasoning

process ensures traceability, reduces error propagation, and mirrors human practices of reflective meeting analysis.

5. Coordination, Communication, and Reasoning Strategies

Effective multi-agent systems rely not only on task decomposition but also on carefully designed coordination, communication, and reasoning mechanisms that govern how agents interact and converge on reliable outcomes. In transcript-based action item extraction, agents typically operate on interdependent subtasks, making structured information exchange essential. Prior research on knowledge-driven multi-agent frameworks demonstrates that shared memory spaces, blackboard architectures, and intermediate symbolic representations enable agents to maintain consistency while reasoning over complex, evolving information (Jin et al., 2025). Such shared representations allow agents to externalize partial interpretations—such as detected intents or candidate actions—so that other agents can inspect, refine, or challenge them.

Iterative feedback loops further enhance reasoning robustness. In this design, downstream agents responsible for validation or consolidation can request re-analysis from upstream comprehension or intent-detection agents when ambiguities or conflicts arise. This mirrors established practices in qualitative analysis, where multiple passes over data are used to improve interpretive reliability (Xu et al., 2025). For meeting transcripts, this iterative coordination is particularly valuable because commitments are often implicit, distributed across turns, or negotiated incrementally through dialogue. By allowing agents to revisit earlier assumptions, the system reduces premature commitments and mitigates cascading errors.

Recent advances in proactive conversational AI suggest that multi-agent systems can go beyond reactive extraction by anticipating follow-up needs and contextual relevance (Deng et al., 2025). In action item extraction, this translates into agents that reason about organizational goals, project context, or prior meetings to refine which actions are most salient. Overall, controlled interaction and distributed reasoning help multi-agent frameworks overcome key limitations of single-model pipelines, including error propagation, opaque decision-making, and limited adaptability to complex conversational dynamics.

6. Applications and Organizational Impact

Multi-agent action item extraction systems have broad applicability across organizational and institutional contexts where meetings serve as a primary coordination mechanism. In corporate environments, such systems support project management by transforming conversational commitments into structured tasks that can be directly integrated into workflow and accountability tools. This capability aligns closely with observed trends in artificial intelligence adoption within professional service industries, where the value of AI is closely tied to its ability to augment decision-making and operational efficiency rather than merely automate documentation (Yang et al., 2024).

In academic and research settings, action item extraction aids collaborative planning, supervision meetings, and interdisciplinary projects by ensuring that agreed-upon responsibilities are explicitly captured and tracked. Customer support and service operations also benefit, as agent-based transcript analysis can identify escalation actions, follow-ups, or promised resolutions embedded within lengthy

interactions. In high-stakes or time-critical environments, such as emergency response training or clinical coordination, accurately extracted action items enhance situational awareness and post-event evaluation.

Visualization and feedback systems built on top of extracted action items further amplify organizational impact. Research on voice record visualization shows that making conversational outcomes explicit improves team communication, learning, and performance, particularly under stress (Oppermann et al., 2025). At the same time, agent-based transcript processing supports privacy- and information-preserving design principles by enabling selective extraction of actionable content rather than wholesale exposure of sensitive conversational data, an approach emphasized in recent work on generative agents for clinical documentation (van Velzen et al., 2025).

7. Challenges and Research Gaps

Despite their promise, multi-agent approaches face several challenges. Transcript quality remains a critical dependency, as errors introduced by automated transcription can propagate through agent pipelines. Implicit commitments and culturally specific communication styles also pose difficulties for intent detection agents, echoing long-standing concerns in conversational and qualitative research (Oliver et al., 2005).

Scalability and evaluation present additional challenges. While multi-agent systems improve reasoning quality, they introduce coordination overhead and complexity. Standardized benchmarks for action item extraction are still limited, although recent work on meeting transcript scarcity and synthetic data generation offers potential pathways forward (Kirstein et al., 2025).

8. Conclusion

Multi-agent systems offer a robust and conceptually well-aligned approach for extracting action items from meeting transcripts by distributing the complex task of conversational understanding across specialized, cooperating agents. Unlike traditional single-model summarization pipelines, multi-agent frameworks enable deeper reasoning about intent, commitment, and context, which are essential for accurately identifying actionable outcomes embedded within unstructured dialogue. By separating comprehension, intent detection, validation, and consolidation into distinct yet coordinated processes, these systems reduce ambiguity, improve interpretability, and enhance the reliability of extracted action items.

This review highlights that the true strength of multi-agent architectures lies not only in task decomposition but also in their coordination and reasoning strategies, which mirror human analytical practices such as iterative review, cross-checking, and contextual refinement. When applied to meeting transcripts, such capabilities support clearer accountability, more effective follow-up, and stronger alignment between conversational decisions and organizational execution. As conversational data continues to grow in scale and importance across professional, academic, and high-stakes domains, multi-agent action item extraction is positioned to become a foundational component of next-generation meeting intelligence and document understanding systems, bridging the gap between discussion and decisive action.

Reference:

- Jin, D., Sun, W., Huang, Jiangping, Liang, P., Xuan, J., Liu, Yang, & Jin, Z. (2025). iReDev: A Knowledge-Driven Multi-Agent Framework for Intelligent Requirements Development (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2507.13081
- Huang, Z., Zhuge, Jinjing, & Zhuge, Jianwei. (2025). Multi-Agent Framework Utilizing Large Language Models for Solving Capture-the-Flag Challenges in Cybersecurity Competitions. Applied Sciences, 15(13), 7159. https://doi.org/10.3390/app15137159
- Xu, Q., Amjad, N., Giles, G., Cumming, A., Hermesky, D., Wen, A., Kwak, M. J., & Kim, Yejin. (2025). A Multi-Agent Large Language Model Framework for Automated Qualitative Analysis (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2512.16063
- Kirstein, F., Khan, Muneeb, Wahle, J. P., Ruas, T., & Gipp, B. (2025). You need to MIMIC to get FAME: Solving Meeting Transcript Scarcity with a Multi-Agent Conversations (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2502.13001
- van Velzen, M., van der Willigen, R. F., de Beer, V. J., de Graaf-Waar, H. I., Janssen, E. R. C., van Leeuwen, S., van der Willigen, M. F., van der Willigen, M. J., Renardus, G., El Maaroufi, R., Satimin, S. J., Hartog, L. M., Hulsen, T., van Meeteren, N. L. U., & Scheper, M. C. (2025). Privacy-, linguistic-, and information-preserving synthesis of clinical documentation through generative agents. Frontiers in Artificial Intelligence, 8. https://doi.org/10.3389/frai.2025.1644084
- Patamia, R. A., Dinh, H. P. T., Liu, M., & Cosgun, A. (2025). Turn-Taking Modelling in Conversational Systems: A Review of Recent Advances. Technologies, 13(12), 591. https://doi.org/10.3390/technologies13120591
- Oliver, D. G., Serovich, J. M., & Mason, T. L. (2005). Constraints and Opportunities with Interview Transcription: Towards Reflection in Qualitative Research. Social Forces, 84(2), 1273–1289. https://doi.org/10.1353/sof.2006.0023
- Yang, J., Blount, Y., & Amrollahi, A. (2024). Artificial intelligence adoption in a professional service industry: A multiple case study. Technological Forecasting and Social Change, 201, 123251. https://doi.org/10.1016/j.techfore.2024.123251
- Oppermann, M., Uhl, J. C., Regal, G., Tscheligi, M., & Murtinger, M. (2025). ROGER: Visualizing Voice Records to Enhance Team Communication Trainings for High-Stress Situations. In Proceedings of the 18th International Symposium on Visual Information Communication and Interaction (pp. 1–9). ACM. VINCI 2025: Proceedings of the 18th International Symposium on Visual Information Communication and Interaction. https://doi.org/10.1145/3769534.3769554
- Moore, R. J. (2015). Automated Transcription and Conversation Analysis. Research on Language and Social Interaction, 48(3), 253–270. https://doi.org/10.1080/08351813.2015.1058600
- Deng, Y., Liao, L., Lei, W., Yang, G. H., Lam, W., & Chua, T.-S. (2025). Proactive Conversational AI: A Comprehensive Survey of Advancements and Opportunities. ACM Transactions on Information Systems, 43(3), 1–45. https://doi.org/10.1145/3715097