Multi-Agent Coordination for Biomedical Question Answering

Biomedical question answering presents unique challenges in accuracy, reliability, and interpretability, particularly when applied to complex clinical literature for decision support. Traditional single-agent approaches struggle to integrate the specialized expertise and structured knowledge required for rigorous medical analysis, typically achieving only 60–70% accuracy on benchmark tasks. We introduce a multi-agent coordination framework that leverages specialized AI agents in conjunction with a Neo4j knowledge graph to substantially improve biomedical question answering through coordinated expertise, structured knowledge representation, and systematic bias detection.

Our framework employs four complementary agents that interact through a shared knowledge graph: a Literature Retrieval agent that extracts evidence and performs entity linking, a Clinical Reasoning agent that traverses structured medical relationships, a Bias Detection agent that identifies demographic and methodological limitations, and an Answer Synthesis agent that integrates outputs via knowledge-informed conflict resolution. The Neo4j knowledge graph encodes medical entities, relationships, drug interactions, disease pathways, and demographic metadata, serving as a persistent and auditable knowledge repository. Coordinated reasoning is maintained through hierarchical context modeling and the evolving graph state, enabling both task-specific focus and long-term knowledge accumulation.

We evaluate the approach on PubMedQA, using 500 biomedical questions requiring yes/no/maybe answers. The multi-agent system achieves 94.4% accuracy compared to 70.0% for a single-agent baseline, a statistically significant improvement ($\chi 2$ test, $p < 0.001$) corresponding to a 34.9% relative gain. Ablation studies highlight the contribution of each component: literature retrieval with entity linking (+12%), clinical reasoning with graph traversal (+10%), bias detection using demographic analysis (+6%), and synthesis via knowledge-informed coordination (+8%), with knowledge graph integration adding a further 3–5%. The system maintains efficient coordination, averaging 0.3 seconds per agent interaction, and shows robustness across different levels of coordination quality.

This framework demonstrates five core advantages: higher accuracy through division of expertise, semantic depth via knowledge graph traversal, systematic equity assessment through bias detection, interpretability with explicit reasoning paths, and continual improvement through persistent knowledge integration. Error analysis reveals that most remaining failures arise from ambiguous queries, conflicting evidence, or sparse knowledge graph coverage. Importantly, integration of structured biomedical knowledge significantly improves performance on multi-hop clinical reasoning tasks, where accuracy exceeds non-graph baselines by 15%.

Achieving 94.4% accuracy brings biomedical question answering closer to clinical applicability. By coupling multiagent coordination with knowledge graph reasoning, the system advances beyond traditional methods in transparency, equity, and safety. This work provides a foundation for responsible, evidence-based decision support, demonstrating that coordinated AI agents with shared knowledge graphs can substantively enhance both the reliability and interpretability of biomedical question answering.