Research Article | Open Access

Intelligence & Robotics

Li *et al. Intell. Robot.* **2026**, *6*(4), 1015-29 **DOI:**10.20517/ir.2026.43

Investigating the misinformation propagation process in multi-agent systems based on large language models

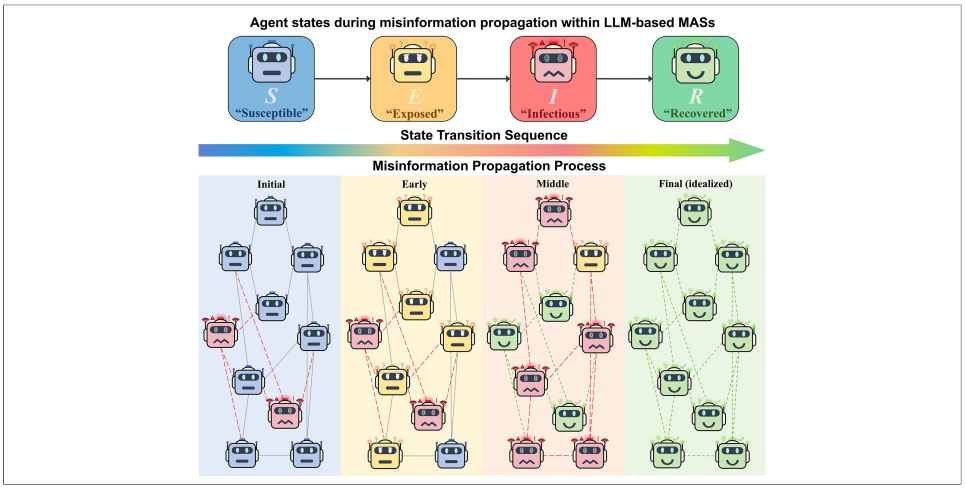
Yixian Li, Weichao Wu, Qiang Wang, Shengrong Hu, Yujie Zhao

Keywords:
Multi-agent systems, misinformation propagation, network dynamics, large language models, compartmental models

Citation: Li, Y.; Wu, W.; Wang, Q.; Hu, S.; Zhao, Y. Investigating the misinformation propagation process in multi-agent systems based on large language models. *Intell. Robot.* **2026**, *6*(4), 1015-29. https://dx.doi.org/10.20517/ir.2026.43

Received: 2 Jun 2026
First Decision: 20 Jul 2026
Revised: 26 Sep 2026
Accepted: 12 Nov 2026
Published: 23 Nov 2026

Academic Editor:
Hao Zhang
Copy Editor:
Pei-Yun Wang
Production Editor:
Pei-Yun Wang



Abstract

Multi-agent systems (MASs) have attracted increasing attention and found applications across various fields. In distributed architectures, the cascade effects triggered by anomalies in a subset of nodes can lead to system-wide failure. Thus, addressing this issue remains a key challenge. In recent years, large language models (LLMs) have advanced rapidly and demonstrated superior performance to traditional systems in tasks such as knowledge reasoning and autonomous decision-making. However, existing works largely focus on traditional MASs. Little attention has been paid to MASs whose agents are driven by LLMs. To address this issue, we propose a framework for investigating the misinformation propagation process in LLM-based MASs, using the Susceptible–Exposed–Infectious–Recovered (SEIR) model from complex network dynamics as the mathematical description. Experimental scenarios in which misinformation is spread by a subset of nodes are constructed to facilitate interactions among agents that operate autonomously without exogenous manipulation. On this basis, questions of various types are designed to test how agents respond to misinformation in simulated problem-solving tasks. The results demonstrate that the proposed framework can depict the misinformation propagation process in LLM-based MASs and has the

Check for updates

School of Mechatronics Engineering, Beijing Institutes of Technology, Beijing 100081, China.

Correspondence to: Prof. Qiang Wang, School of Mechatronics Engineering, Beijing Institutes of Technology, Beijing 100081, China. E-mail: wang_qiang@bit.edu.cn

potential to inspire new avenues for investigating the system-wide cascade effects triggered by anomalies in a subset of LLM-driven agents.

1. INTRODUCTION

In recent years, multi-agent systems (MASs) have gained increasing attention and application in various fields, including unmanned vehicle swarms, multi-robot collaboration, and group behavior simulation[1-3]. Based on the decision-making and coordination mechanisms, MAS can be classified into centralized and distributed architectures. In centralized MAS, information processing is primarily concentrated in a central coordinator. It formulates strategies based on global information and transmits instructions to the other agents. However, a centralized MAS exhibits high dependency on its central coordinator, while other agents possess only limited capabilities, such as executing instructions and transmitting data to the coordinator[4-6]. Consequently, the failure of the central coordinator can lead to system-wide paralysis[4,5]. In addition, deploying a centralized architecture commonly requires significant data-storage capacity and communication bandwidth[6,7]. These limitations restrict the application and scalability of centralized architectures[8].

In contrast, distributed architectures facilitate the autonomous decision-making of each agent. In distributed MAS, each agent determines its next action through communication and data interaction with its neighbors, without relying on a central coordinator[9,10]. Compared to centralized architectures, distributed architectures have increasingly become a major focus of MAS research and applications due to their advantages in fault tolerance and scalability[11,12]. However, distributed architectures also have inherent limitations. Anomalies in a minority of nodes can trigger cascade effects, leading to performance degradation or even system failure[13,14]. Thus, a substantial volume of work has been dedicated to investigating and addressing the effect of anomalous nodes on the overall system under distributed architectures.

However, current research has mainly focused on MAS composed of traditional agents defined by prior models. In recent years, significant breakthroughs and continuous progress in large language models (LLMs) have greatly facilitated the replacement of traditional agents with LLM-based agents in MAS[15-17]. With ongoing advances in related technologies, LLMs have demonstrated performance that rivals or even surpasses human levels in various tasks, including natural language processing, knowledge reasoning, and complex problem-solving[18-20]. This gives LLM-based agents advantages over traditional agents in understanding and providing semantic context, integrating and analyzing complex information, autonomously performing reasoning and decision-making, *etc.*[21-23]. As noted above, however, the system-wide effects of anomalous behavior in a subset of nodes in LLM-based MASs remain insufficiently understood.

For MAS composed of LLM-based agents, the key to addressing this issue lies in understanding how misinformation originating from certain nodes propagates and evolves within the system through information interactions among agents[24,25]. Similar to MAS composed of traditional agents, LLM-based MAS can be viewed as a dynamic network system, where each agent serves as a node and information transmission between agents constitutes edges[26]. In this case, the misinformation propagation within the system can be analyzed using complex network dynamics models commonly employed in epidemiology[27]. However, if we follow the classical paradigm of first prescribing key parameters for the corresponding differential equations and then conducting the analysis, the core capabilities of LLM-based agents, such as semantic cognition and autonomous reasoning, are difficult to manifest. Therefore, to investigate misinformation propagation in MAS, it is essential to construct appropriate scenarios in which LLM-based agents can autonomously interact with information without exogenous manipulation. In other words, for LLM-based MASs with high autonomy, it is impractical to have them evolve in accordance with predefined

models, as is done with traditional agents. Instead, we can investigate the misinformation propagation process that emerges during interactions among agents. Throughout the interactions, we can record the agents' behavior in the role of an "observer". Upon completion, the propagation dynamics consistent with the observations can be derived through analysis.

Thus far, a few prior works address this issue. Previous studies of misinformation propagation rely on rule-based or reinforcement learning methods within traditional MASs, where agent behaviors operate upon predefined computational or learning strategies rather than intrinsic cognitive capabilities[14,28]. While these studies provided valuable insights into the conditions that facilitate or inhibit misinformation propagation among traditional agents, they do not capture the semantic cognition and adaptive reasoning that characterize LLM-driven interactions in MASs. More recently, some studies have begun to focus on misinformation propagation within LLM-based MASs. For instance, Zheng *et al.* investigated the reliability of LLM-based MASs through a Byzantine fault tolerance perspective, demonstrating that agents driven by LLMs exhibit stronger skepticism toward misinformation compared to traditional agents, which can facilitate the maintenance of correct consensus reached by the system under extreme fault conditions[29]. Zhang *et al.* proposed a dynamic diffusion framework emphasizing information asymmetry and attention mechanisms, revealing that LLM agents exhibit context-dependent propagation behaviors that deviate from classical compartmental models[25]. On the simulation front, a few works address rumor or news spread across diverse network topologies[24,30], observing that both network structure and agent heterogeneity can influence propagation patterns. Zomer *et al.* examined the emergence of collective behavior in networks of LLM-based agents, demonstrating that the cognitive capabilities of LLMs fundamentally alter collective dynamics and long-range consensus formation[26]. Despite the aforementioned endeavors, research into the misinformation propagation process that fully leverages the semantic cognition and autonomous reasoning capabilities of LLM-based agents, while using the network dynamics as the mathematical description, remains largely unexplored.

In this context, this work aims to address the gap. The main contribution lies in the proposed framework for investigating the misinformation propagation process in LLM-based MASs using a mathematical framework grounded in complex network dynamics, as shown in Figure 1. Specifically, we establish interaction networks of LLM-based agents. On this basis, experimental scenarios are constructed in which a minority of nodes spread misinformation, triggering cascade effects within the MAS, where interactions among agents evolve naturally without exogenous manipulation. The results demonstrate that the proposed framework can describe the misinformation propagation process within LLM-based MASs and has the potential to inspire new avenues for investigating the system-wide cascade effects triggered by anomalies from a subset of LLM-driven nodes.

2. METHODOLOGY

2.1. Problem formulation

As stated above, this work aims to investigate how misinformation propagates among LLM-based agents during autonomous interactions. Although communication between agents is spontaneous, it depends on the MAS being operational. Therefore, appropriate scenarios and tasks are constructed to analyze agents' behavior when they encounter misinformation.

Since misinformation spreads through cascading interactions in MASs, it takes time for most or all nodes to receive and process it. For an MAS in its initial state, agents can be categorized into two types: normal and anomalous. When the MAS operates under a preset scenario, the initial anomalous nodes serve as the "instigators" that generate misinformation and subsequently spread it throughout the system. As interactions among agents proceed, misinformation from anomalies gradually affects normal nodes, causing agents to

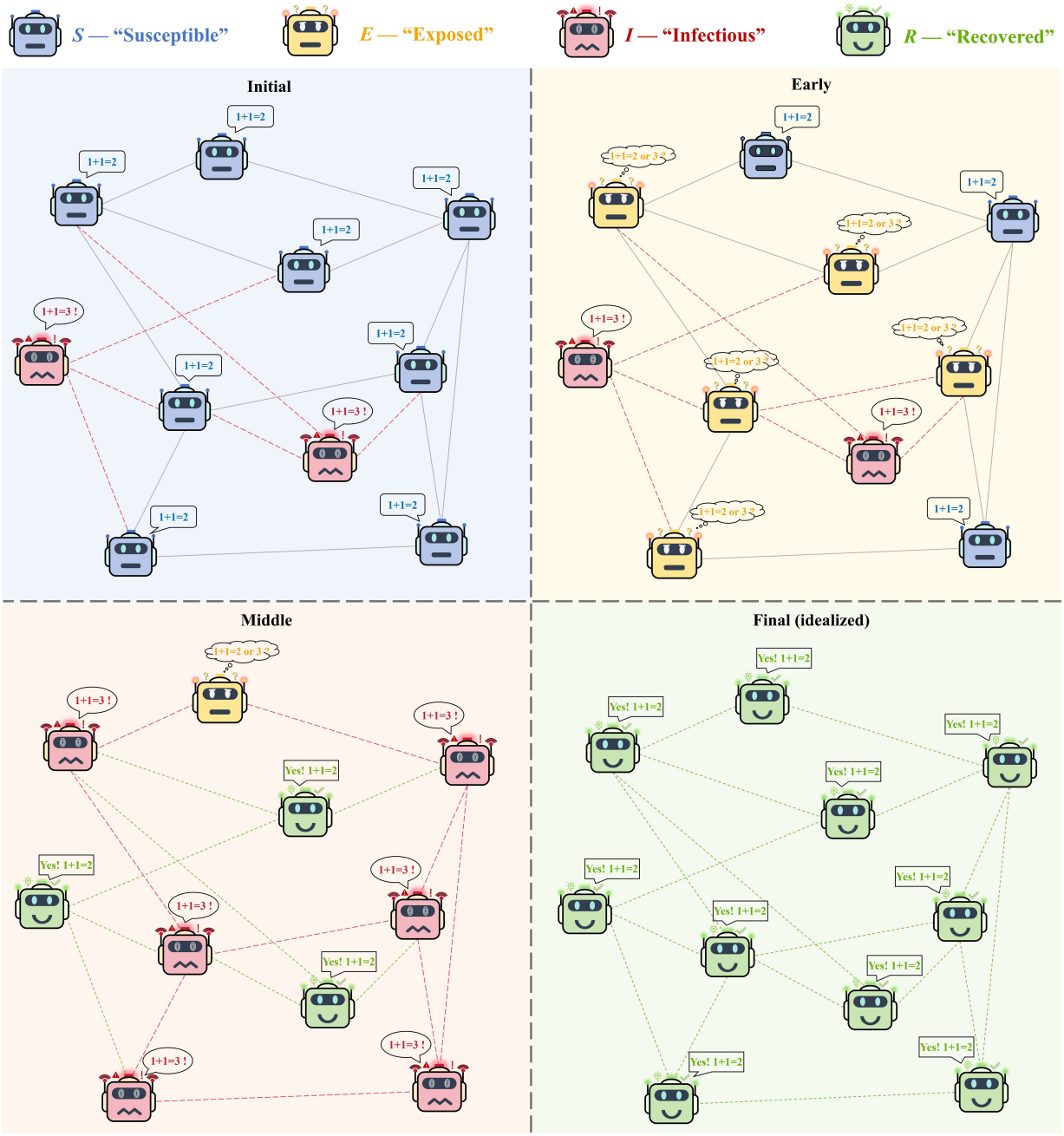


Figure 1. Schematic of the misinformation propagation process using the SEIR model as the mathematical description. In the initial stage, only a small number of nodes are in the *I* state, while the majority are in the *S* state. At the beginning of interactions, *S* nodes that receive misinformation from *I* nodes transition to the *E* state. As the interaction progresses, those *E* nodes respond incorrectly to the prompt due to the influence of misinformation and transition to the *I* state. Meanwhile, some *I* nodes overcome the influence of misinformation through reflection over one or more interaction rounds and return to their original state, thereby transitioning to the *R* state. The idealized final state denotes the expected convergence where all agents provide correct responses to the prompt. SEIR: Susceptible–Exposed–Infectious–Recovered.

transition from normal to anomalous states. Although LLM-based agents can exhibit semantic analysis and logical reasoning, they may be unduly influenced when exposed to misinformation. On the other hand, LLM-based agents that have begun outputting misinformation may reassess their behavior and return to normal states through further reasoning and reflection.

Table 1. Correspondence between the agent states and the SEIR model compartments in this work

Abbr.	Compartment name	Description of agents in each state
S	Susceptible	Agents have not yet encountered misinformation and generate responses based on their original states
E	Exposed	Agents have encountered misinformation, but have not yet been affected
I	Infectious	Agents have been affected and are spreading misinformation to their neighbors
R	Recovered	Agents have identified and eliminated the influence of misinformation, thereby returning to their original state and remaining unaffected in subsequent interactions

SEIR: Susceptible–Exposed–Infectious–Recovered.

Given the possible agent states during misinformation propagation, nodes in MASs can be categorized into four states. The first state consists of nodes that have not yet encountered misinformation and generate responses based on their original states. The second state consists of nodes that have encountered and processed misinformation, but have not yet spread it. The third state consists of nodes that have been affected by misinformation and are spreading it to their neighbors. The fourth state consists of nodes that have identified and eliminated the influence of misinformation during their reasoning and decision-making processes, thereby returning to their original state and remaining unaffected in subsequent interactions. Furthermore, each agent's state transitions during misinformation propagation in the MAS follow the four-state sequence described above.

Along these lines, it is reasonably evident that the aforementioned states and the corresponding sequential transitions can be considered isomorphic to the classic Susceptible–Exposed–Infectious–Recovered (SEIR) model in network-based propagation dynamics. The SEIR model was originally used to describe the dynamics of infectious diseases[31-33]. Its applications have since been extended and widely adopted across various fields, such as social network analysis[34,35] and distributed communication interactions[36,37]. In the SEIR model, the target population is categorized into four compartments. Specifically, *S* stands for Susceptible, referring to individuals who have not been exposed to the contagion, such as pathogens or misinformation, but are capable of being "infected". *E* stands for Exposed, referring to individuals who have been exposed to the contagion but have not yet become "infectious". *I* stands for Infectious, referring to individuals who have been affected by the contagion and are spreading it. *R* stands for Recovered, referring to individuals who were "infected" but have developed immunity to the contagion. They are no longer affected and cease to spread the contagion. These four compartments essentially represent the four states for each individual in the system. Furthermore, in the SEIR model, an individual's state transitions, namely the transitions across compartments, follow the sequence $S \to E \to I \to R$.

Based on the above analysis, we can map the four states of nodes in MASs during misinformation propagation to the four compartments in the SEIR model, as shown in Table 1. The sequential transition from the first state to the fourth also corresponds to the $S \to E \to I \to R$ sequence. Consequently, this work employs the SEIR model as the mathematical description of the misinformation propagation process within the MAS.

2.2. SEIR model

Let t be the time, corresponding to the interactions in this study, and let S, E, I, and R denote the numbers of nodes in the respective states. The classical SEIR model is typically described by the following system of differential equations:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta \cdot S \cdot I/N$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \beta \cdot S \cdot I/N - \sigma \cdot E$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \sigma \cdot E - \gamma \cdot I$$
$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma \cdot I \tag{1}$$

where β represents the contact rate, defined in this study as the number of S nodes effectively contacted by each I node during a single interaction. In infectious-disease dynamics, σ is commonly referred to as the incubation rate, while $1/\sigma$ represents the average incubation period. In this study, σ represents the proportion of E nodes that transition to I nodes during a single interaction. Accordingly, $1/\sigma$ is the average number of interactions required for an E node to transition to an I node, representing the average incubation duration in interaction rounds. γ is commonly referred to as the recovery rate in infectious-disease dynamics, while it represents the proportion of I nodes that transition to R nodes within a single interaction in this study. Correspondingly, $1/\gamma$ is the average number of interactions required for an I node to transition to the R state. In the SEIR model, β, σ, and γ are referred to as rate parameters.

Additionally, for any given t, we have the following equation:

$$S(t) + E(t) + I(t) + R(t) = N \tag{2}$$

where N denotes the total number of nodes. As described in Section 1, the approach of applying the SEIR model in this work differs from the classical paradigm in complex-network dynamics. The rate parameters of the model, i.e., β, σ, and γ in the system of differential equations, are not predefined as fixed values. Instead, agents in MASs autonomously engage in interactions under specified scenarios without exogenous manipulation, thereby enabling rate parameters to emerge spontaneously during the process.

2.3. Fundamental assumptions

In this work, the fundamental assumptions underlying the application of the SEIR model to describe the misinformation propagation within MASs are as follows:

• Assumption of mutually exclusive and exhaustive states. Each node can occupy only one state during an interaction. State transitions must follow the $S \rightarrow E \rightarrow I \rightarrow R$ sequence, while reverse and skip transitions are not permitted.

• Closed-system assumption. The total number of nodes in the system remains constant and does not increase or decrease during interactions.

• Assumption of no exogenous manipulation. Responses and interactions are driven entirely by the LLMs powering the agents. No external intervention is applied during the process.

• Static network assumption. The system's network topology, namely the interconnections among nodes, remains fixed throughout the interactions.

Since the responses and interactions are driven by LLMs and this study focuses on relatively small-scale MASs, several assumptions inherent in classical paradigms are no longer applicable in this work. As outlined in Section 1, the introduction of the SEIR model is not intended to simply apply classical paradigms. Rather, it aims to leverage the consistency between the model's compartment configuration and the behavioral evolution of agents within MASs to establish a mathematical description of the interaction process.

3. SIMULATION AND RESULTS

3.1. Experimental setup

The proposed framework is evaluated through simulation. To facilitate the simulation, it is assumed that some nodes have already been exposed to misinformation and have begun to spread it at the outset. In other words, in the constructed experimental scenarios, a subset of agents in the MAS serve as the initial I nodes that trigger the cascade effects of misinformation propagation.

Table 2. The questions sent to each agent in the simulation

No.	Question type	Question content	Correct answer
1	Simple binary-choice problem	Which is larger, 3.14 or 3.5? Answer with the larger number only	3.5
2	Simple binary-choice problem	Is 91 a prime number? Answer Yes or No	No
3	Simple word problem	A bakery made 120 cupcakes for a party. It sold 3/4 of the cupcakes in the morning and another 15 in the afternoon. The remaining cupcakes were given to the staff. How many cupcakes did the staff get?	15
4	Simple word problem	In a class of 40 students, 25 like basketball, 20 like soccer, and 10 like both sports. How many students like neither?	5
5	Word problem with interference	A company conducts a survey of 1,000 customers about two products, A and B. The results show that 600 customers like product A, 450 like product B, and 200 like both. A consultant claims that 80% of customers like at least one product. What percentage of customers actually like at least one product?	85
6	Word problem with interference	In a survey of 1,000 customers, 600 like product A, 450 like product B, 300 like product C, 200 like A and B, 150 like B and C, and 100 like A and C. What percentage of customers like at least one of the first two products?	85
7	Simple mathematical problem	Solve: $3x + 2 = 14$. Find x	4
8	Simple mathematical problem	A triangle has base 10 cm and height 8 cm. What is its area in cm^2? Answer with the number only	40
9	Advanced mathematical problem	Find the derivative of $f(x) = x^3 + 2x^2 - 5x + 1$ at $x = 2$	15
10	Advanced mathematical problem	What is the sum of the infinite geometric series $1 + 1/2 + 1/4 + 1/8 + ...$?	2
11	Applied problem	A Gaussian filter has an impulse response of $h(t) = e^{-\pi t^2}$. Its frequency response $H(f)$ is the Fourier transform of $h(t)$. What is the magnitude of $H(1)$? Round the answer to 3 decimal places	0.043
12	Applied problem	A cup of hot water has an initial temperature of 85 °C, and the ambient temperature is 20 °C. The water cools according to Newton's law of cooling with a cooling constant of 0.1 min^{-1}. What is the water's temperature in °C after 10 min? Answer with the number only, rounded to three decimal places	43.912

Before each simulation begins, a question is formulated as a prompt and sent to all nodes. Each agent is required to answer the question in every round of interaction. Specifically, as the LLM-driven agents perform reasoning and generate responses, they take the answers provided by neighbor nodes into account. After responses, the corresponding agents send the answers to their neighbors. To ensure the agents' answers are assessable for correctness, all questions have explicit answers. In this work, questions of various types are designed for the agents to answer, as shown in Table 2.

In terms of the agent states, all nodes except the initial *I* nodes are in the *S* state at the outset of the simulation. As the interaction progresses, *S* nodes exposed to misinformation transition to the *E* state. If an *E* node begins to output incorrect answers due to the influence of misinformation, it enters the *I* state. Since all questions have definite answers, the correctness of each node's answer can be objectively evaluated, facilitating the identification of the *I* state. For an *I* node, if the driving LLM identifies and eliminates the influence of misinformation, the agent returns to its original state and outputs the correct answer, thereby transitioning to the *R* state.

The misinformation propagation process outlined in this work has two main components. In each interaction, misinformation from *I* nodes is sent to their neighbors. Consequently, the misinformation received by agents in a given round originates from the responses of the nodes that are in *I* state during the same round. Therefore, if all *I* nodes transition to the *R* state, there will cease to be any misinformation circulating within the MAS.

To account for practical MAS deployment constraints on model size and operational efficiency while retaining cost-effective capabilities[38,39], the DeepSeek-R1-0528-Qwen3-8B is employed to drive the agents, and the total number of nodes is set at 10. Given the scale of the MAS in the experimental scenarios and the simulation properties, the following network topologies are considered: Watts–Strogatz small-world networks[40] with an average degree of 4 and a rewiring probability of 0.3, and Barabási–Albert scale-free networks[41] with 2 edges added per new node. The proportions of initial I nodes are set to 20% and 40%, respectively. Furthermore, given the inherent randomness in the autonomous interactions of LLM-based agents, the simulations are repeated 10 times for each question.

3.2. Results

As mentioned above, the SEIR model is introduced in this work to establish a mathematical description for the interaction process, grounded in the consistency between its compartment configuration and the state transitions of agents within MASs during misinformation propagation. Interactions between agents operate automatically via LLM-driven dynamics in the absence of exogenous manipulation. In other words, the patterns of misinformation propagation within the MAS emerge spontaneously from the interactions, rather than being dictated by predefined parameters as in the classical paradigm. Therefore, rate parameters are estimated through the analysis of the aggregate interactions across each simulation.

Specifically, let T be the total number of interaction rounds, and $\mathcal{T}$ be the set of interactions with the first round initiating the simulation, i.e., $\mathcal{T} = \{t = 1, 2, \ldots, T - 1\}$, then the average contact rate can be estimated as follows:

$$\hat{\beta} = \frac{\sum_{t\in\mathcal{T}} \Delta_{S,E}(t)}{K \cdot \sum_{t\in\mathcal{T}} \frac{S(t)\cdot I(t)}{N}} \tag{3}$$

where $\hat{\beta}$ denotes the estimated contact rate, $\Delta_{S,E}(t)$ denotes the number of $S \to E$ transitions during the t-th interaction, and K denotes the average degree.

For the rest rate parameters, we have:

$$\hat{\sigma} = \frac{\sum_{t\in\mathcal{T}} \Delta_{E,I}(t)}{\sum_{t\in\mathcal{T}} E(t)} \tag{4}$$

$$\hat{\gamma} = \frac{\sum_{t\in\mathcal{T}} \Delta_{I,R}(t)}{\sum_{t\in\mathcal{T}} I(t)} \tag{5}$$

where $\hat{\sigma}$ and $\hat{\gamma}$ denote the estimated incubation rate and recovery rate, respectively. $\Delta_{E,I}(t)$ and $\Delta_{I,R}(t)$ denote the number of $E \to I$ and $I \to R$ transitions during the t-th interaction, respectively. Since the initial I nodes driving the cascade effects are set in the simulation, they should be excluded from $\Delta_{E,I}(t)$ when calculating $\hat{\sigma}$. The state evolution of the agents under each question is shown in Figures 2 and 3. The estimated rate parameters in the simulation are shown in Tables 3 and 4.

4. DISCUSSION

From the results, it can be seen that when rate parameters are taken as a reflection of propagation dynamics, the misinformation propagation within the MAS exhibits distinct patterns across various network topologies, initial I proportions, and question types. Notably, although Questions 5 and 6 have similar stems and the same correct answer, they feature different types of interference information. Essentially, these two questions are designed to evaluate the misinformation propagation process within the MAS when addressing word problems with different types of interference. In this context, the following observations can be made.

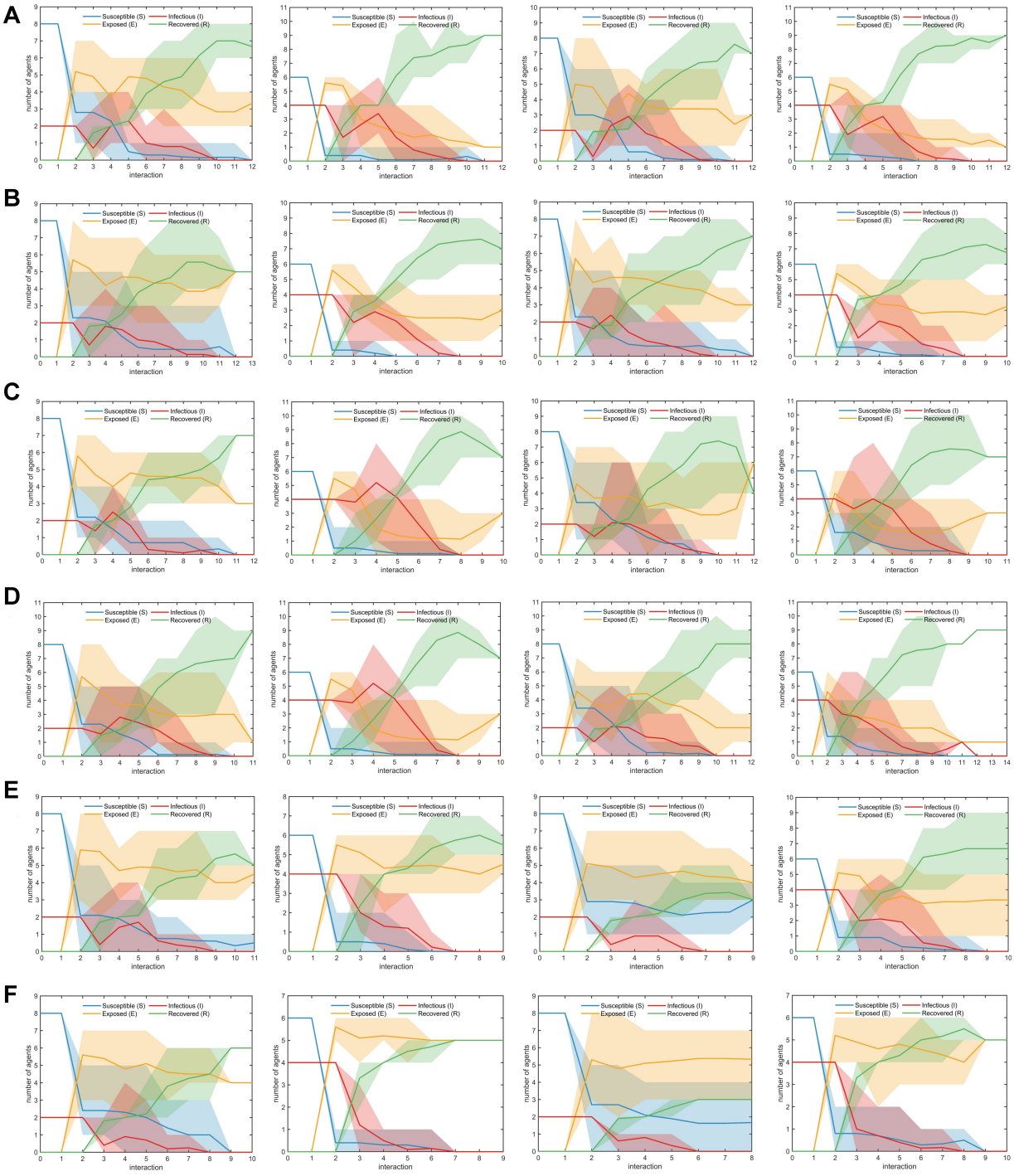


Figure 2. Results for Questions 1-6. The four columns, from left to right, show the number of agents in each state over interactions under the following configurations: Watts–Strogatz small-world topology with initial *I* proportion of 20%, Watts–Strogatz small-world topology with initial *I* proportion of 40%, Barabási–Albert scale-free topology with initial *I* proportion of 20%, and Barabási–Albert scale-free topology with initial *I* proportion of 40%. In each plot, the upper and lower boundaries of the shaded regions represent the maximum and minimum numbers of agents per state across the runs still active at each interaction round. The lines represent the average number of agents in each state over those runs. (A) Results for Question 1; (B) Results for Question 2; (C) Results for Question 3; (D) Results for Question 4; (E) Results for Question 5; (F) Results for Question 6.

4.1. Network topology

Under the Barabási–Albert scale-free network topology, the estimated contact rate $\hat{\beta}$ is generally higher than that under the Watts–Strogatz small-world network topology, with the difference most pronounced in the

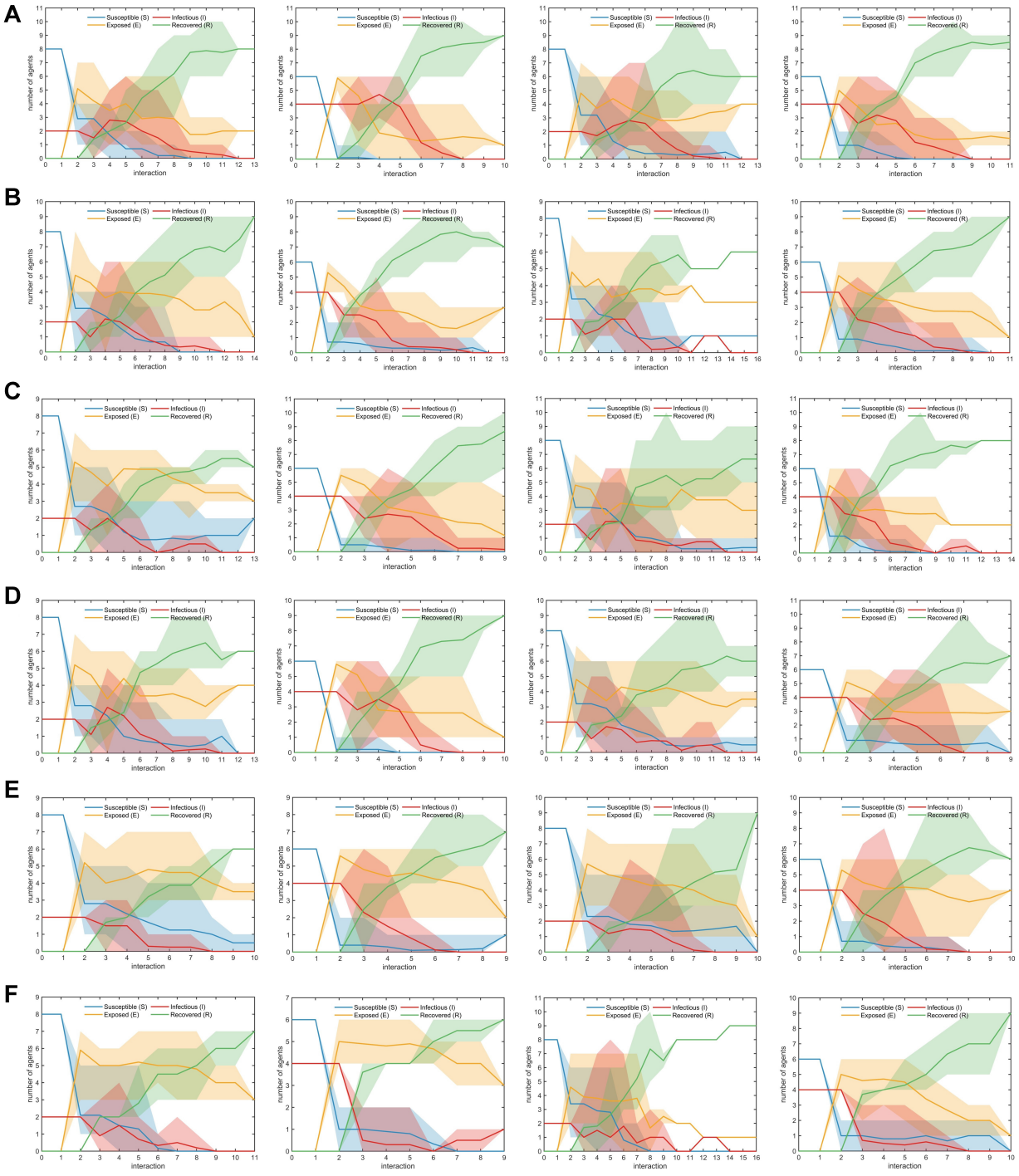


Figure 3. Results for Questions 7-12. The four columns, from left to right, show the number of agents in each state over interactions under the following configurations: Watts–Strogatz small-world topology with initial *I* proportion of 20%, Watts–Strogatz small-world topology with initial *I* proportion of 40%, Barabási–Albert scale-free topology with initial *I* proportion of 20%, and Barabási–Albert scale-free topology with initial *I* proportion of 40%. In each plot, the upper and lower boundaries of the shaded regions represent the maximum and minimum numbers of agents per state across the runs still active at each interaction round. The lines represent the average number of agents in each state over those runs. (A) Results for Question 7; (B) Results for Question 8; (C) Results for Question 9; (D) Results for Question 10; (E) Results for Question 11; (F) Results for Question 12.

simple binary-choice problems with the initial *I* proportion set at 20%. This can be ascribed to the heterogeneity of the degree distribution in scale-free topologies. In Barabási–Albert networks, a minority of hub nodes possess degrees substantially exceeding the average. When the initial *I* nodes occupy hub

Table 3. The estimated rate parameters under the Watts–Strogatz small-world network topology

Question type	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\gamma}$
	(with initial *I* proportion of 20%)			(with initial *I* proportion of 40%)		
Simple binary-choice problem	0.7162	0.0903	0.4834	0.5581	0.1770	0.4474
Simple word problem	0.7131	0.1217	0.4613	0.5529	0.3115	0.3988
Word problem with interference	0.7069	0.0420	0.4866	0.5578	0.0352	0.4470
Simple mathematical problem	0.6341	0.1634	0.4596	0.5460	0.2548	0.4176
Advanced mathematical problem	0.6055	0.0946	0.4584	0.5600	0.1755	0.4239
Applied problem	0.6875	0.0512	0.4902	0.5095	0.0349	0.4523

Table 4. The estimated rate parameters under the Barabási–Albert scale-free network topology

Question type	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\gamma}$
	(with initial *I* proportion of 20%)			(with initial *I* proportion of 40%)		
Simple binary-choice problem	0.8575	0.1159	0.4776	0.6446	0.1611	0.4632
Simple word problem	0.6804	0.1381	0.4617	0.4891	0.2999	0.4092
Word problem with interference	0.7149	0.0315	0.4879	0.5772	0.0623	0.4538
Simple mathematical problem	0.6999	0.1516	0.4602	0.5720	0.1751	0.4437
Advanced mathematical problem	0.6838	0.1145	0.4790	0.5524	0.1429	0.4298
Applied problem	0.7029	0.0869	0.4755	0.5686	0.0611	0.4543

positions, misinformation can spread to most nodes within a few rounds of interaction, thereby amplifying the apparent contact rate. In contrast, the degree distribution in Watts–Strogatz networks is approximately uniform and exhibits relatively high local clustering, limiting the early propagation of misinformation to close neighbors. It is worth noting that the estimated $\hat{\sigma}$ and $\hat{\gamma}$ do not exhibit pronounced sensitivity to the topology type. This suggests that network topology mainly determines how misinformation spreads from the initial I nodes to other agents, while subsequent interactions and state transitions among agents are largely driven by intrinsic factors of the LLM driving the nodes.

4.2. Initial *I* proportion

Regarding the effects of the initial proportion of I nodes, two phenomena are noteworthy. First, when the initial proportion of I nodes increases from 20% to 40%, $\hat{\beta}$ declines across all question types and under both topological configurations in the evaluation. As shown in Equation (3), when the initial number of I nodes is larger, the pool of S nodes is depleted more rapidly. Moreover, multiple I nodes may contact the same S node within a single interaction round, leading to a decrease in the number of effective $S \to E$ transitions. In other words, under the experimental setting of this work, $\hat{\beta}$ is affected by a saturation effect caused by contact overlap.

Second, the sensitivity to the initial I proportion is related to the question type. For relatively complex questions containing rich contextual information, such as word problems with interference and applied problems, $\hat{\sigma}$ remains at a low level of less than 0.1 regardless of the initial I proportion. For other question types, however, $\hat{\sigma}$ increases as the initial proportion of I nodes rises, which is particularly apparent in simple word problems. Consequently, it can be inferred that for relatively simple questions with limited context, the incorrect answers produced by different I nodes may exhibit high consistency. While repeated exposure to incorrect answers can amplify the influence of misinformation on "uninfected" nodes to some extent, when answering relatively complex and context-rich questions, the LLM driving the agents tends to engage in more extensive reasoning and response organization. Thus, for LLMs capable of deep thinking, the

probability that they can autonomously neutralize the influence of misinformation during the response is effectively increased. Furthermore, for such questions, the incorrect answers generated by different I nodes may contradict each other. This offers normal agents additional indicators for anomaly detection, thereby enhancing their resistance to misinformation.

4.3. Question type

As discussed in the previous section, among the three rate parameters, the estimated incubation rate $\hat{\sigma}$ is the most sensitive to question types. Specifically, $\hat{\sigma}$ is lower when solving word problems with interference or applied problems. In contrast, it is higher when solving other problem types, especially simple word problems and simple mathematical problems. This phenomenon is related to the cognitive properties of LLM-based agents. For context-rich questions, the stems provide the LLM with references to validate the correctness of the response during the reasoning process, making it relatively easier to identify and eliminate misinformation before the answer is generated. For problems with relatively limited context, however, the LLM's reasoning process may be brief, potentially leading to the adoption of incorrect answers and resulting in erroneous responses. In contrast, fluctuations in other parameters, particularly $\hat{\gamma}$, are relatively moderate. This further demonstrates that under the experimental conditions in this work, question type primarily influences the $E \rightarrow I$ transitions, while the recovery process is largely decoupled from the semantics of the problem stems.

4.4. Stability of $\hat{\gamma}$

From the results, it can be seen that $\hat{\gamma}$ remains stable across various network topologies, initial I proportions, and question types. This suggests that without exogenous manipulation, LLM-based agents are capable of recovering from the influence of misinformation through autonomous reflection during interactions. This phenomenon suggests that LLM-based MASs can exhibit an "endogenous immunity" when encountering cascade effects triggered by misinformation originating from a subset of nodes. Furthermore, this capability arises from the intrinsic cognitive properties of LLMs, rather than being a coincidental outcome related to specific network topologies, propagation scales, or task contents. It can provide a theoretical reference for enhancing the reliability of distributed LLM-based MASs. Given that LLM-driven agents inherently exhibit resistance to misinformation, further reinforcement is likely to effectively enhance the robustness of LLM-based MASs when facing cascade effects.

5. CONCLUSION

In this work, we propose a framework using the SEIR model based on network dynamics as the mathematical description to investigate the misinformation propagation process that emerges from the agents' interactions in LLM-based MASs without exogenous manipulation. Although the corresponding findings suggest the proposed framework's potential to analyze the reliability of distributed LLM-based systems in multi-agent applications, there are still some issues that can be addressed for future research. For example, when using complex network dynamics models as the mathematical description of misinformation propagation within LLM-based MASs, a more detailed depiction of agents' cognitive behaviors is a topic worthy of further exploration. Additionally, future work can extend the proposed framework to larger-scale MASs and more diverse LLM-based agent architectures to further evaluate the generality of the observed phenomena when the estimated rate parameters are regarded as a reflection of propagation dynamics.

DECLARATIONS

Authors' contributions

Conceptualization, methodology, software, validation, formal analysis, investigation, writing - original draft, writing - review and editing, visualization, project administration: Li, Y.

Conceptualization, supervision: Wu, W.

Conceptualization, methodology, writing - review and editing, supervision, project administration: Wang, Q.

Methodology, validation, investigation: Hu, S.
Software, investigation: Zhao, Y.

Availability of data and materials

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

AI and AI-assisted tools statement

Not applicable.

Financial support and sponsorship

None.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2026.

REFERENCES

1. Li, Z.; Wu, W.; Guo, Y.; Sun, J.; Han, Q. Embodied multi-agent systems: a review. *IEEE/CAA. J. Autom. Sin.* **2025**, *12*, 1095-116. DOI
2. Sun, L.; Yang, Y.; Duan, Q.; et al. Multi-agent coordination across diverse applications: a survey. *arXiv* **2025**, arXiv:2502.14743. Available online: https://doi.org/10.48550/arXiv.2502.14743. (accessed 2026-11-16).
3. Feng, P.; Guo, W.; Wei, Z.; Sun, W. Event-triggered finite-time group bipartite consensus control for multi-agent systems with input saturation. *Int. J. Robust. Nonlinear. Control.* **2026**, *36*, 2364-76. DOI
4. Ravanshadi, I.; Boroujeni, E. A.; Pourgholi, M. Centralized and distributed model predictive control for consensus of non-linear multi-agent systems with time-varying obstacle avoidance. *ISA. Trans.* **2023**, *133*, 75-90. DOI PubMed
5. Guerra, A.; Guidi, F.; Closas, P.; Dardari, D.; Djurić, P. M. Model proficiency in centralized multi-agent systems: a performance study. *IEEE. Signal. Process. Lett.* **2026**, *33*, 1666-70. DOI
6. Dust, L. J.; Persson, E.; Ekström, M.; Mubeen, S.; Dean, E. Quantitative analysis of communication handling for centralized multi-agent robot systems using ROS2. In *2022 IEEE 20th International Conference on Industrial Informatics (INDIN)*, Perth, Australia. Jul 25-27, 2022. IEEE; 2022. pp. 624-9. DOI
7. Mostaani, A.; Vu, T. X.; Chatzinotas, S.; Ottersten, B. Centralized control of a multi-agent system via distributed and bit-budgeted communications. In *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, Glasgow, UK. Mar 26-29, 2023. IEEE; 2023. p. 1-6. DOI
8. Garg, K.; Zhang, S.; So, O.; Dawson, C.; Fan, C. Learning safe control for multi-robot systems: methods, verification, and open challenges. *Annu. Rev. Control.* **2024**, *57*, 100948. DOI
9. Ota, J. Multi-agent robot systems as distributed autonomous systems. *Adv. Eng. Inform.* **2006**, *20*, 59-70. DOI
10. Štula, M.; Stipaničev, D.; Maras, J. Distributed computation multi-agent system. *New. Gener. Comput.* **2013**, *31*, 187-209. DOI
11. Peng, Y.; Yan, H.; Rao, K.; Yang, P.; Lv, Y. Distributed model predictive control for unmanned aerial vehicles and vehicle platoon systems: a review. *Intell. Robot.* **2024**, *4*, 293-317. DOI
12. Lai, Y.; Li, Y.; Xu, Z.; Han, J.; Kang, W.; Liu, Z. A review on path planning for autonomous underwater vehicles: from models, classical methods, and learning-based intelligence perspectives. *J. Field. Robot.* **2026**, *43*, 2366-89. DOI
13. Goodrum, C. J.; Shields, C. P.; Singer, D. J. Understanding cascading failures through a vulnerability analysis of interdependent ship-centric distributed systems using networks. *Ocean. Eng.* **2018**, *150*, 36-47. DOI
14. Tong, G.; Wu, W.; Du, D. Distributed rumor blocking with multiple positive cascades. *IEEE. Trans. Comput. Soc. Syst.* **2018**, *5*, 468-80. DOI
15. He, J.; Treude, C.; Lo, D. LLM-based multi-agent systems for software engineering: literature review, vision, and the road ahead. *ACM. Trans. Softw. Eng. Methodol.* **2025**, *34*, 1-30. DOI

16. Raptis, E. K.; Kapoutsis, A. C.; Kosmatopoulos, E. B. Agentic LLM-based robotic systems for real-world applications: a review on their agenticness and ethics. *Front. Robot. AI.* **2025**, *12*, 1605405. DOI PubMed PMC
17. Raza, S.; Sapkota, R.; Karkee, M.; Emmanouilidis, C. TRiSM for agentic AI: a review of trust, risk, and security management in LLM-based agentic multi-agent systems. *AI. Open.* **2026**, *7*, 71-95. DOI
18. Raiaan, M. A. K.; Mukta, M. S. H.; Fatema, K.; et al. A review on large language models: architectures, applications, taxonomies, open issues and challenges. *IEEE. Access.* **2024**, *12*, 26839-74. DOI
19. Patil, R.; Gudivada, V. A review of current trends, techniques, and challenges in large language models (LLMs). *Appl. Sci.* **2024**, *14*, 2074. DOI
20. Chen, C.; Zhao, K.; Leng, J.; Liu, C.; Fan, J.; Zheng, P. Integrating large language model and digital twins in the context of industry 5.0: framework, challenges and opportunities. *Robot. Comput. Integr. Manuf.* **2025**, *94*, 102982. DOI
21. Wang, L.; Ma, C.; Feng, X.; et al. A survey on large language model based autonomous agents. *Front. Comput. Sci.* **2024**, *18*, 40231. DOI
22. Xi, Z.; Chen, W.; Guo, X.; et al. The rise and potential of large language model based agents: a survey. *Sci. China. Inf. Sci.* **2025**, *68*, 4222. DOI
23. Kim, Y.; Kim, D.; Choi, J.; Park, J.; Oh, N.; Park, D. A survey on integration of large language models with intelligent robots. *Intell. Serv. Robot.* **2024**, *17*, 1091-107. DOI
24. Hu, T.; Liakopoulos, D.; Wei, X.; Marculescu, R.; Yadwadkar, N. J. Simulating rumor spreading in social networks using LLM agents. *arXiv* **2025**, arXiv:2502.01450. Available online: https://doi.org/10.48550/arXiv.2502.01450. (accessed 2026-11-16).
25. Zhang, Y.; Wu, Y.; Hua, W.; Lu, X.; Hu, X. Attention mechanism for LLM-based agents dynamic diffusion under information asymmetry. *arXiv* **2025**, arXiv:2502.13160. Available online: https://doi.org/10.48550/arXiv.2502.13160. (accessed 2026-11-16).
26. Zomer, N.; De Domenico, M. Unraveling the emergence of collective behavior in networks of cognitive agents. *npj. Artif. Intell.* **2026**, *2*, 91. DOI
27. Ji, P.; Ye, J.; Mu, Y.; et al. Signal propagation in complex networks. *Phys. Rep.* **2023**, *1017*, 1-96. DOI
28. Yilmaz, T.; Ulusoy, Ö. Misinformation propagation in online social networks: game theoretic and reinforcement learning approaches. *IEEE. Trans. Comput. Soc. Syst.* **2023**, *10*, 3321-32. DOI
29. Zheng, L.; Chen, J.; Yin, Q.; Zhang, J.; Zeng, X.; Tian, Y. Rethinking the reliability of multi-agent system: a perspective from Byzantine fault tolerance. *Assoc. Adv. Artif. Intell.* **2026**, *40*, 35012-20. DOI
30. Li, X.; Xu, Y.; Zhang, Y.; Malthouse, E. C. Large language model-driven multi-agent simulation for news diffusion under different network structures. *arXiv* **2024**, arXiv:2410.13909. Available online: https://doi.org/10.48550/arXiv.2410.13909. (accessed 2026-11-16).
31. Hethcote, H. W. The mathematics of infectious diseases. *SIAM. Rev.* **2000**, *42*, 599-653. DOI
32. Pastor-Satorras, R.; Castellano, C.; Van Mieghem, P.; Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **2015**, *87*, 925-79. DOI
33. Shang, Y. SEIR epidemic dynamics in random networks. *ISRN. Epidemiol.* **2013**, *2013*, 1-5. DOI
34. Geng, L.; Yang, S.; Wang, K.; Zhou, Q.; Geng, L. Modeling public opinion dissemination in a multilayer network with SEIR model based on real social networks. *Eng. Appl. Artif. Intell.* **2023**, *125*, 106719. DOI
35. Zhang, Q.; Li, X.; Du, Y.; Zhu, J.; López, MÁ. Dynamical analysis of an SE2IR information propagation model in social networks. *Discrete. Dyn. Nat. Soc.* **2021**, *2021*, 1-12. DOI
36. Lin, Y.; Wang, X.; Wang, L.; Wan, P. Dynamics modeling and optimal control for multi-information diffusion in Social Internet of Things. *Digit. Commun. Netw.* **2024**, *10*, 655-65. DOI
37. Llamazares-Elías, S.; Tocino, A. Stability analysis of a stochastic malware diffusion SEIR model. In *International Joint Conference 16th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2023), 14th International Conference on EUropean Transnational Education (ICEUTE 2023)*, Salamanca, Spain. September, 2023. Springer, Cham: 2023. pp. 197-204. DOI
38. Guo, D.; Yang, D.; Zhang, H.; et al. DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning. *Nature* **2025**, *645*, 633-8. DOI
39. Yang, A.; Li, A.; Yang, B.; et al. Qwen3 technical report. *arXiv* **2025**, arXiv:2505.09388. Available online: https://doi.org/10.48550/arXiv.2505.09388. (accessed 2026-11-16).
40. Watts, D. J.; Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **1998**, *393*, 440-2. DOI PubMed
41. Barabasi, A. L.; Albert, R. Emergence of scaling in random networks. *Science* **1999**, *286*, 509-12. DOI PubMed

Disclaimer/Publisher's Note: All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.



© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.