

# Revisão Sistemática da Literatura integrando Sistemas Multiagente e *Large Language Models*<sup>\*</sup>

Míriam B. Born<sup>1</sup>, Bruno C. Alves<sup>1</sup>, Felipe M. Goulart<sup>1</sup>,  
Leticia B. Caldas<sup>1</sup> e Marilton S. de Aguiar<sup>1</sup>

<sup>1</sup>Universidade Federal de Pelotas (UFPel)  
Pelotas – RS – Brasil

{mbborn, bcalves, fmgoulart, lbcaldas, marilton}@inf.ufpel.edu.br

**Abstract.** *The systematic literature review investigates studies conducted within a specific field. It aims to focus on a well-defined topic to identify, select, evaluate, and synthesize evidence related to a specific research question. The purpose of a systematic literature review is to provide a comprehensive overview of a problem, offering a standardized and detailed account of the scientific studies. This study presents a systematic review of two areas: Multi-Agent Systems (MAS) and Large Language Models (LLMs), considering that the integration of these fields can contribute to addressing real-world problems across various domains of knowledge.*

**Resumo.** *A revisão sistemática da literatura investiga os estudos realizados em uma determinada área. Desta forma, busca focar em um tema bem definido com intuito de identificar, selecionar, avaliar e sintetizar evidências de um tópico de pesquisa específico. A revisão sistemática da literatura possui como objetivo proporcionar uma visão ampla sobre um problema, com um detalhamento padronizado sobre os estudos científicos. Este estudo apresenta uma revisão sistemática de duas áreas: os Sistemas Multiagente (SMA) e os Large Language Models (LLMs), visto que a integração dessas áreas pode contribuir em demandas de problemas reais em diversas áreas do conhecimento.*

## 1. Introdução

A área de Inteligência Artificial (IA) está se tornando cada vez mais relevante devido à sua implicação direta na sociedade. Essa área tem como objetivo principal reproduzir habilidades racionais humanas por meio de ferramentas computacionais [Russell and Norvig 2020]. As aplicações de inteligência artificial podem ser implementadas em diversas áreas, como agricultura, educação e saúde.

Uma das abordagens mais promissoras dentro da IA é o uso de Sistemas Multiagente (SMA). De acordo com [Wooldridge 2009], esse tipo de sistema é composto por múltiplos agentes capazes de realizarem ações para satisfazerem seus objetivos. Desta maneira, os agentes tomam decisões para resolver problemas complexos, de forma distribuída e colaborativa, dentro de um ambiente específico ao longo do tempo.

---

\*O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Com o avanço de modelos de IA, surgem novas possibilidades para integração de sistemas multiagentes, como os *Large Language Models* (LLMs). Segundo [Ozdemir 2023], LLMs são modelos de aprendizado profundo treinados com grandes volumes de dados textuais para compreender e gerar linguagem humana, como GPT-3 [Brown et al. 2020]. Assim, sistemas multiagentes baseados em LLMs buscam aprimorar a inteligência coletiva e as habilidades especializadas dos agentes.

A Revisão Sistemática da Literatura (RSL) consiste em um método científico específico que vai um passo além da simples visão geral. Uma RSL visa identificar, selecionar, avaliar, interpretar e sintetizar os principais estudos disponíveis em uma área com base em uma questão de pesquisa. Cada RSL envolve objetivos específicos, o que permite aos pesquisadores analisarem criticamente os dados coletados, a fim de resumir evidências existentes sobre um tópico de interesse, para identificar e posicionar novas possibilidades de pesquisa ainda não investigadas [Kitchenham and Charters 2007].

O objetivo deste estudo, diante dessas premissas, é apresentar uma revisão sistemática da literatura, referente aos últimos cinco anos, sobre a integração entre as áreas de Sistemas Multiagente e *Large Language Models*. Portanto, a partir desta RSL, será possível identificar padrões, desafios e oportunidades de pesquisa envolvendo SMA e LLM.

O artigo está estruturado da seguinte forma: na Seção 2 são apresentados os principais conceitos das áreas de Sistemas Multiagente e *Large Language Models*. A Seção 3 apresenta a metodologia aplicada à Revisão Sistemática da Literatura neste artigo, como base de dados e palavras-chave utilizadas. A Seção 4 mostra os artigos resultantes da RSL e a abordagem proposta por cada um. Por fim, na Seção 5 são apresentadas as conclusões e os trabalhos futuros do artigo.

## 2. Referencial Teórico

Esta Seção apresenta os conceitos sobre Agentes e Sistemas Multiagente, bem como sobre os *Large Language Models*.

### 2.1. Agentes e Sistemas Multiagente

Sistemas Multiagente (SMA) são compostos por múltiplas entidades autônomas, denominadas agentes, que interagem em um ambiente compartilhado para alcançar objetivos individuais ou coletivos. Cada agente possui capacidades específicas, como percepção, raciocínio e ação, permitindo-lhe tomar decisões de forma independente. A interação entre esses agentes pode resultar em comportamentos emergentes complexos, tornando os SMA particularmente úteis na modelagem de sistemas distribuídos e dinâmicos, como redes de computadores, sistemas de transporte e simulações sociais [Born et al. 2023].

Os agentes em SMA podem ser classificados em duas categorias principais: reativos e cognitivos. Agentes reativos respondem diretamente a estímulos do ambiente, sem representação interna ou memória de ações passadas, sendo eficazes em sistemas que requerem respostas rápidas a mudanças ambientais. Por outro lado, agentes cognitivos possuem representações internas do ambiente e dos outros agentes, permitindo-lhes planejar, aprender com experiências passadas e tomar decisões mais complexas [Alvares and Sichman 1997].

A arquitetura de um SMA define como os agentes são organizados e interagem. Arquiteturas simples podem envolver agentes homogêneos com funções semelhantes, enquanto arquiteturas complexas podem incluir agentes heterogêneos com diferentes papéis e níveis de autonomia. As interações entre agentes podem ser diretas, por meio de troca de mensagens, ou indiretas, através de um ambiente compartilhado. A aplicação de SMA é vasta, abrangendo áreas como robótica, onde múltiplos robôs colaboram para realizar tarefas; sistemas de transporte, para coordenação de veículos autônomos; e simulações sociais, para estudar comportamentos emergentes em sociedades humanas. A capacidade dos SMA de lidar com complexidade, adaptabilidade e distribuição de tarefas os torna uma abordagem poderosa para resolver problemas em ambientes dinâmicos e incertos [Girardi 2004].

## **2.2. Large Language Models**

Modelos de Linguagem de Grande Escala (LLMs, do inglês *Large Language Models*), são sistemas avançados de inteligência artificial projetados para compreender e gerar linguagem humana de forma natural. Esses modelos são treinados em vastas quantidades de dados textuais, utilizando técnicas de aprendizado profundo, como as redes neurais do tipo *transformer*. Essa arquitetura permite que os LLMs capturem padrões complexos da linguagem, possibilitando-lhes realizar tarefas como tradução automática, resumo de textos, resposta a perguntas e geração de conteúdo coerente e contextualmente relevante [Thirunavukarasu et al. 2023]. São modelos generativos treinados para estimar a distribuição estatística de *tokens* de textos gerados por humanos, onde os *tokens* em questão incluem palavras, partes de palavras ou caracteres individuais — inclusive pontuações [Shanahan 2024].

A capacidade dos LLMs de gerar texto natural os torna ferramentas valiosas em diversas aplicações, desde assistentes virtuais até sistemas de recomendação e análise de sentimentos. Sua versatilidade e adaptabilidade decorrem do treinamento em conjuntos de dados diversificados, que abrangem diferentes estilos, tópicos e contextos linguísticos. Com o avanço contínuo da tecnologia, espera-se que os LLMs desempenhem um papel cada vez mais significativo na automação de tarefas cognitivas e na facilitação da interação entre humanos e máquinas [Shanahan 2024].

## **3. Metodologia**

Nesta Seção, é descrito o protocolo utilizado para a realização desta Revisão Sistemática da Literatura, tendo como foco responder à seguinte questão de pesquisa: “Quais são os principais estudos envolvendo Sistemas Multiagente com integração a *Large Language Models*?”.

A Tabela 1 sumariza os principais elementos envolvidos no protocolo seguido nesta RSL, compreendendo a questão principal de pesquisa, as questões específicas, o objetivo da revisão, os critérios de inclusão e os critérios de exclusão de estudos. Além disso, foram definidas as bases de pesquisa, as quais estão descritas através da Tabela 2. O processo de triagem e revisão dos artigos foi realizado pelos quatro autores deste estudo, de forma colaborativa, com decisões tomadas por consenso entre os revisores para garantir maior confiabilidade na seleção dos artigos.

**Tabela 1. Protocolo da Revisão Sistemática da Literatura.**

<b>Questão principal</b>	“Quais são os principais estudos envolvendo sistemas multiagente com integração a Large Language Models?”
<b>Questões específicas</b>	i) “Quais são as áreas de aplicação para os sistemas multiagente?”; ii) “Quais são os LLMs utilizados nos estudos?”; e, iii) “Quais são as ferramentas consideradas para simulação dos sistemas multiagente?”.
<b>Objetivo</b>	O objetivo desta revisão sistemática é encontrar estudos na literatura, em especial dos últimos cinco anos (2020-2025), sobre Sistemas Multiagente integrados com <i>Large Language Models</i> .
<b>Critérios de inclusão</b>	Artigos publicados que contenham, no título ou resumo, as palavras-chave ou seus sinônimos: Sistema Multiagentes e <i>Large Language Models</i> .
<b>Critérios de exclusão</b>	i) Estudos não escritos em inglês; ii) Estudos não encontrados nas bases definidas; iii) Estudos que não contenham as palavras-chave; iv) Livros completos, artigos resumidos, estudos secundários, duplicados ou redundantes; e, v) Estudos que não tratam do tema central de pesquisa.

**Tabela 2. Bases de dados utilizadas neste estudo.**

Nome	URL
ACM	<a href="https://dl.acm.org/">https://dl.acm.org/</a>
ScienceDirect	<a href="https://www.sciencedirect.com">https://www.sciencedirect.com</a>

Considerando a questão principal de pesquisa e o objetivo deste trabalho, foram definidas as palavras-chave e a *string* de busca, as quais podem ser observadas na Tabela 3. Com o protocolo, a base de dados e a *string* de busca definidos, foi realizada a pesquisa por trabalhos nas fontes de dados. Esse processo de busca foi realizado em abril de 2025 e compreende trabalhos dos últimos cinco anos.

**Tabela 3. Termos utilizados para definir a *string* de busca da RSL.**

(“multi-agent systems” <b>OR</b> “multiagent systems” <b>OR</b> “multi-agent system” <b>OR</b> “multiagent system”)
<b>AND</b>
(“large language models” <b>OR</b> “large language model” <b>OR</b> “LLM” <b>OR</b> “LLMs” <b>OR</b> “LLM-based”)

Foram encontrados 40 trabalhos nas bases de dados (27 em ACM e 13 em ScienceDirect). Com isso, foi realizada a pesquisa das informações básicas dos artigos no formato *bibtex*. Para auxílio durante o processo de revisão dos artigos, utilizou-se a ferramenta de software *Parsifal*<sup>1</sup>, a qual permite gerenciar os artigos, realizar anotações e atribuir *tags*. Além disso, o *Parsifal* pode buscar por artigos duplicados de forma automática. Entretanto, neste momento do estudo não foram encontrados estudos duplicados.

Inicialmente, foi realizada a fase de avaliação dos estudos em que foram analisados cada um dos artigos coletados através de leitura do título e resumo. Nesta fase, verifica-se se os dois componentes dos artigos estão relacionados com a questão principal de pesquisa. Essa fase é importante para eliminar artigos com base nos critérios de exclusão. Por exemplo, um artigo que apresenta a aplicação de agentes inteligentes em

<sup>1</sup><https://parsif.al/>

logística, mas não menciona ou integra nenhum LLM, deve ser excluído. Após a etapa de avaliação do título e resumo, aprovaram-se 15 artigos.

Por fim, foi realizada a fase de avaliação em que analisou-se cada um dos artigos coletados através da leitura completa. Nesta etapa, verificou-se quais estudos melhor descrevem as questões de pesquisa da RSL e atendem ao objetivo desta revisão. Deste modo, foram aprovados 12 artigos, conforme estão apresentados no Seção 4, mais precisamente sumarizados na Tabela 4.

#### 4. Resultados e Discussões

Nesta Seção são apresentados os resultados deste estudo. Deste modo, são discutidas as abordagens propostas pelos trabalhos selecionados durante esta RSL.

O estudo de [Wang et al. 2025b] apresenta um sistema composto por diversos agentes especializados para melhorar o reconhecimento de entidades em textos, sem a necessidade de dados anotados. Para a construção do sistema foi utilizado GPT-3.5 como LLM. Os autores destacam que, ao dividir a tarefa entre agentes que avaliam contextos, escolhem bons exemplos e refinam as previsões, o desempenho melhora significativamente, mesmo sem supervisão tradicional.

No estudo de [Chen et al. 2024] os autores propuseram uma estrutura chamada BlockAgents com a integração da tecnologia de *blockchain* em um SMA cooperativo baseado em LLM (GPT-4 e Bard) com intuito de mitigar comportamentos bizantinos. Os ataques bizantinos são inspirados no *Problema dos Generais Bizantinos*, que neste contexto do estudo alguns agentes do sistema comportam-se de maneira arbitrária. A estrutura desenvolvida neste trabalho segue um fluxo bem definido onde são inseridos: atribuição de funções, declaração de proposta, avaliação e tomada de decisão. A partir dessa metodologia foram propostos os experimentos em três conjuntos de dados os quais mostram que o BlockAgents reduz a interferência de ataques de envenenamento na precisão para menos de 3% e reduz a taxa de sucesso de ataques de *backdoor* para menos de 5%, demonstrando a capacidade de resistência contra ataques bizantinos.

No artigo de [Ryu et al. 2025], os autores propõem e avaliam um sistema interativo no qual usuários conversam com múltiplos agentes da área cinematográfica. Neste sistema são criadas dinâmicas sociais simuladas, onde os usuários podem escolher pessoas relacionadas ao cinema para interagir, como personagens, diretores e fãs. Deste modo, o foco deste estudo está em entender como essas conversas influenciam a forma como as pessoas interpretam e valorizam os filmes, usando GPT-4o.

O trabalho de [Yue et al. 2024] apresenta um SMA, construído em Python, para automatização de tarefas no contexto de ensaios clínicos. O sistema faz uso do modelo GPT-4, sendo capaz de interpretar linguagem natural, integrar informações de bases biomédicas e apoiar decisões clínicas. Desta forma, é capaz de supervisionar ensaios clínicos por meio de interações conversacionais e geração de resultados, combinando agentes baseados em LLM com modelos de aprendizado de máquina.

Dango, sistema proposto por [Chen et al. 2025], busca aprimorar a comunicação entre agentes baseados em LLMs e usuários por meio de linguagem natural. Deste modo, o sistema tem como objetivo melhorar a clareza de intenção do usuário, sendo utilizado para isto o GPT-4o-mini como parte do SMA. Assim, a ferramenta pode auxiliar usuários

reais a realizar tarefas complexas, como a de preparação de dados.

O artigo proposto por [Fatemi and Hu 2024] apresenta um sistema onde múltiplos agentes especializados colaboram entre si para analisar diversos tipos de dados financeiros. Estes agentes fazem uso de LLM, mais especificamente GPT-4o-mini. Neste contexto, a aplicação é capaz de lidar com dados multimodais, como texto e gráficos, além de fornecer explicações transparentes sobre o raciocínio dos agentes.

Na pesquisa de [Li et al. 2025] foi desenvolvida uma abordagem com intuito de mitigar um problema de jogos cooperativos em sua fase inicial. Esta metodologia é composta por aprendizado por reforço multiagente (MARL) e LLMs, sendo que esses modelos fornecem estratégias no início do treinamento. Neste trabalho os autores propuseram um conjunto de ferramentas abrangendo algumas funcionalidades como: um ambiente integrado SMAC (*StarCraft Multi-Agent Challenge*), a geração automática de *prompts* em Processamento de Linguagem Natural, visualização e treinamento de algoritmos de comunicação entre agentes, interface de avaliação e foram utilizados os modelos Qwen2.5-72b, Qwen-max, LlaMA3-8b, LlaMA3-70b, GLM-4, GPT-4, DeepSeek-r1-8b e DeepSeek-r1-70b.

[Shi et al. 2025] propõem um *framework* multiagente para reduzir alucinações em LLMs durante tarefas de raciocínio complexo. Deste modo, o objetivo do sistema é melhorar a precisão e a estabilidade das respostas de LLMs, por meio de diálogos, oferecendo um modelo prático para integrar agentes com filtragem colaborativa e memória adaptativa. Para mitigação das alucinações, foram utilizados diversos LLMs, como GPT-3, GPT-3.5-Turbo e GPT-4-Turbo.

O estudo de [Wang et al. 2025a] apresenta um SMA capaz de integrar LLMs com algoritmos evolutivos para resolver problemas complexos de agendamento em ambientes industriais e logísticos. A arquitetura do sistema é composta por agentes especializados que atuam de forma colaborativa e iterativa, incluindo agentes para definição de problemas, criação de soluções candidatas, otimização e validação. Para isto, os autores utilizaram o *framework* AgentScope para o SMA e o GPT-4o como LLM dos agentes.

Os autores [Orner et al. 2024] apresentam um sistema no qual agentes tomam decisões cooperativas sobre um tópico, a partir de opiniões diversas e interpretações emocionais extraídas por meio de análise de sentimentos. Utilizando o modelo GPT-3.5-Turbo como LLM dos agentes, o sistema busca definir o consenso coletivo, levando em conta que os sentimentos dos agentes mudam ao longo do tempo. A abordagem foi avaliada no contexto de recrutamento de pessoas, no qual os agentes avaliam candidatos.

O trabalho proposto por [Qiao et al. 2025] apresenta um sistema multiagente desenvolvido para automatizar a análise temática, com o objetivo de identificar padrões de significados em grandes volumes de textos não estruturados. Como estudo de caso, os autores utilizaram textos sobre mudanças climáticas, os quais foram codificados, organizados e agregados. Os agentes do sistema foram implementados com o uso do *framework* AutoGen, sendo utilizado o modelo GPT-4o como base dos agentes de LLM.

Por fim, em [Cinkusz and Chudziak 2024] foi desenvolvido o CogniSim que utiliza, em seus principais componentes, a integração de SMA e LLMs com técnicas de Contexto Dinâmico e criação de perfis de agentes, visando aprimorar a colaboração entre estes com a utilização da Teoria da Mente. O modelo de LLM utilizado no estudo

foi o GPT-4 e o CogniSim propõe a análise de problemas, soluções para os mesmos, construção e validação de planos e coordenação de agentes especialistas que podem desempenhar funções de desenvolvedores, executores, agentes focados na qualidade e agentes revisores de metodologia. Os agentes do CogniSim são responsáveis pela produção de documentação, modelos e diagramas UML considerando medidas pré definidas de qualidade e desempenho.

Os estudos selecionados nesta revisão sistemática da literatura apresentam diferentes abordagens, conforme discutido anteriormente. A Tabela 4 sumariza os trabalhos selecionados nesta RSL, destacando-se as áreas de aplicação dos Sistemas Multiagente implementados e os *Large Language Models* utilizados como integração dos respectivos sistemas. Apesar da diversidade de aplicações, áreas do conhecimento e modelos, observa-se que poucos estudos utilizam *frameworks* consolidados de programação multiagente, sendo que todos fazem uso de tecnologias baseadas em GPT. Esse cenário representa oportunidades relevantes para pesquisas futuras.

**Tabela 4. Resultados da Revisão Sistemática da Literatura**

Artigo	Área do Conhecimento	LLM
A Cooperative Multi-Agent Framework for Zero-Shot Named Entity Recognition [Wang et al. 2025b]	Processamento de Linguagem Natural	GPT-3.5
BlockAgents: Towards Byzantine-Robust LLM-Based Multi-Agent Coordination via Blockchain [Chen et al. 2024]	Computação	GPT-4, Bard
Cinema Multiverse Lounge: Enhancing Film Appreciation via Multi-Agent Conversations [Ryu et al. 2025]	Interação Humano-Computador	GPT-4o
ClinicalAgent: Clinical Trial Multi-Agent System with Large Language Model-based Reasoning [Yue et al. 2024]	Saúde	GPT-4
Dango: A Mixed-Initiative Data Wrangling System using Large Language Model [Chen et al. 2025]	Processamento de Linguagem Natural	GPT-4o-mini
FinVision: A Multi-Agent Framework for Stock Market Prediction [Fatemi and Hu 2024]	Finanças	GPT-4o-mini
LLM-guided decision-making toolkit for multi-agent reinforcement learning [Li et al. 2025]	Aprendizado por Reforço Multiagente	LlaMA3-8b, LlaMA3-70b, GLM-4, GPT-4, DeepSeek-r1-8b
Mitigating reasoning hallucination through Multi-agent Collaborative Filtering [Shi et al. 2025]	Computação	GPT-3, GPT-3.5-Turbo, GPT-4-Turbo
Multi-agent large language models as evolutionary optimizers for scheduling optimization [Wang et al. 2025a]	Logística	GPT-4o
Sentimental Agents: Combining Sentiment Analysis and Non-Bayesian Updating for Cooperative Decision-Making [Orner et al. 2024]	Recursos Humanos	GPT-3.5-Turbo
Thematic-LM: A LLM-based Multi-agent System for Large-scale Thematic Analysis [Qiao et al. 2025]	Análise Temática	GPT-4o
Towards LLM-augmented multiagent systems for agile software engineering [Cinkusz and Chudziak 2024]	Engenharia de Software	GPT-4

## 5. Conclusões

Este artigo apresentou uma breve Revisão Sistemática de Literatura considerando o estudo nas áreas de Sistemas Multiagente (SMA) e *Large Language Models* (LLM), a fim de identificar trabalhos relevantes de uma área específica na literatura. Deste modo, neste estudo detalhou-se o procedimento para elaboração de uma RSL, sendo possível de ser reproduzida a partir das informações metodológicas seguidas.

Apesar da área dos grandes modelos de linguagem ser relativamente recente, as pesquisas em Agentes e SMA ocorrem há diversos anos com aplicações nas mais variadas

áreas do conhecimento. Entretanto, a integração dessas duas áreas considerando agentes e SMA autônomos/inteligentes está em ascensão.

Considerando as áreas de interesse deste estudo, foram selecionados 12 trabalhos relevantes na literatura. Nesta pesquisa, constatou-se que os modelos LLMs mais utilizados foram o GPT e a interação entre os agentes dos estudos foram realizadas via *prompt*, considerando em diversos trabalhos os agentes como sistemas/estruturas conversacionais, sem a utilização de uma ferramenta de programação multiagente.

De forma geral, os trabalhos utilizados nesta RSL foram muito satisfatórios, a partir da visão geral do tópico de pesquisa nos últimos cinco anos e da prova que esta área tem potenciais de contribuições. Como trabalhos futuros pretende-se realizar testes com *frameworks* desenvolvidos para a programação multiagente como JaCaMo [Boissier et al. 2016] e GAMA [Taillandier et al. 2019], e também outras LLMs e integrar com ferramentas de programação multiagente.

## Referências

- Alvares, L. H. T. and Sichman, J. S. (1997). Introdução aos sistemas multiagentes. In *Jornada de Atualização em Informática, 16.; Congresso da SBC, 17., Brasília*, pages 1–38. SBC.
- Boissier, O., Hübner, J. F., and Ricci, A. (2016). *The JaCaMo Framework*, pages 125–151. Springer International Publishing, Cham.
- Born, M., Aguiar, M., and Adamatti, D. (2023). Modeling organizational level of agents with moise+ to a natural resources management rpg. In *The international FLAIRS conference proceedings*, volume 36.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language models are few-shot learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS ’20, Red Hook, NY, USA. Curran Associates Inc.
- Chen, B., Li, G., Lin, X., Wang, Z., and Li, J. (2024). Blockagents: Towards byzantine-robust llm-based multi-agent coordination via blockchain. In *Proceedings of the ACM Turing Award Celebration Conference-China 2024*, pages 187–192.
- Chen, W.-H., Tong, W., Case, A., and Zhang, T. (2025). Dango: A mixed-initiative data wrangling system using large language model. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI ’25, New York, NY, USA. Association for Computing Machinery.
- Cinkusz, K. and Chudziak, J. A. (2024). Towards llm-augmented multiagent systems for agile software engineering. In *Proceedings of the 39th IEEE/ACM International Conference on Automated Software Engineering*, pages 2476–2477.
- Fatemi, S. and Hu, Y. (2024). Finvision: A multi-agent framework for stock market prediction. In *Proceedings of the 5th ACM International Conference on AI in Finance*,

- ICAIIF '24, page 582–590, New York, NY, USA. Association for Computing Machinery.
- Girardi, R. (2004). Introdução aos sistemas multiagentes. DevMedia. Disponível em: <https://www.devmmedia.com.br/introducao-aos-sistemas-multiagentes/28973>.
- Kitchenham, B. A. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report, United Kingdom.
- Li, Z., Zhang, R., Wang, Z., Xie, Z., and Song, Y. (2025). Llm-guided decision-making toolkit for multi-agent reinforcement learning. *Neurocomputing*, 638:130105.
- Orner, D., Ondula, E. A., Mumero Mwangi, N., and Goyal, R. (2024). Sentimental agents: Combining sentiment analysis and non-bayesian updating for cooperative decision-making. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '24, page 2408–2410, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.
- Ozdemir, S. (2023). *Quick Start Guide to Large Language Models: Strategies and Best Practices for Using ChatGPT and Other LLMs*. Addison-Wesley, Boston, MA, USA.
- Qiao, T., Walker, C., Cunningham, C., and Koh, Y. S. (2025). Thematic-lm: A llm-based multi-agent system for large-scale thematic analysis. In *Proceedings of the ACM on Web Conference 2025*, WWW '25, page 649–658, New York, NY, USA. Association for Computing Machinery.
- Russell, S. and Norvig, P. (2020). *Artificial Intelligence: A Modern Approach*. Pearson Education, Hoboken, NJ, USA, 4th edition.
- Ryu, J., Kim, K., Heo, D., Song, H., Oh, C., and Suh, B. (2025). Cinema multiverse lounge: Enhancing film appreciation via multi-agent conversations. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA. Association for Computing Machinery.
- Shanahan, M. (2024). Talking about large language models. *Communications of the ACM*, 67(2):68–79.
- Shi, J., Zhao, J., Wu, X., Xu, R., Jiang, Y.-H., and He, L. (2025). Mitigating reasoning hallucination through multi-agent collaborative filtering. *Expert Systems with Applications*, 263:125723.
- Taillandier, P., Gaudou, B., Grignard, A., Quang Nghi, H., Marilleau, N., Caillou, P., Philippin, D., and Drogoul, A. (2019). Building, composing and experimenting complex spatial models with the gama platform. *GeoInformatica*, 23.
- Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., and Ting, D. S. W. (2023). Large language models in medicine. *Nature medicine*, 29(8):1930–1940.
- Wang, Y., Wang, J., and Chu, Z. (2025a). Multi-agent large language models as evolutionary optimizers for scheduling optimization. *Computers Industrial Engineering*, 206:111197.

- Wang, Z., Zhao, Z., Lyu, Y., Chen, Z., de Rijke, M., and Ren, Z. (2025b). A cooperative multi-agent framework for zero-shot named entity recognition. In *Proceedings of the ACM on Web Conference 2025*, WWW '25, page 4183–4195, New York, NY, USA. Association for Computing Machinery.
- Wooldridge, M. (2009). *An Introduction to MultiAgent Systems*. Wiley, Chichester, West Sussex, England, 2nd edition.
- Yue, L., Xing, S., Chen, J., and Fu, T. (2024). Clinicalagent: Clinical trial multi-agent system with large language model-based reasoning. In *Proceedings of the 15th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, BCB '24, New York, NY, USA. Association for Computing Machinery.