

Arquiteturas distribuídas e integração de agentes inteligentes: evolução, desafios e cenário atual dos MAS



Distributed architectures and integration of intelligent agents: evolution, challenges, and the current landscape of MAS

Leonardo Costa Sá 

Fatec Praia Grande
leonardo.c.sa@hotmail.com

Simone Maria Viana Romano 

Fatec Praia Grande
prof.simone.viana@gmail.com

PALAVRAS-CHAVE: Inteligência artificial; automação de sistemas inteligentes; arquitetura de sistemas multiagentes; Model Context Protocol – MCP.

ABSTRACT

The recent advancement of Artificial Intelligence (AI), particularly generative models, has expanded the applications of intelligent systems in distributed environments. In this context, Multi-Agent Systems (MAS) have emerged as a key strategy for coordinating autonomous agents in complex tasks. The main objective of this article is to present the state of the art of MAS from the perspective of contemporary Artificial Intelligence, with particular emphasis on its integration with Large Language Models (LLMs) and the emerging role of the Model Context Protocol (MCP) as a facilitating element in this interaction. The methodology employed is qualitative, grounded in a narrative literature review with a survey of publications in databases such as Google Scholar and Scopus. Additionally, Excel was used for the segmentation and critical analysis of the documents. The findings of this investigation indicate that MAS has become increasingly sophisticated in response to the growing complexity of distributed environments, particularly with the support of generative AI. The integration with LLMs and the emergence of MCP as a communication protocol play a central role in this progress. MCP enables interoperability and distributed autonomous decision-making, showing promise despite its early stage of adoption. Challenges remain, and future research should assess its effectiveness in more diverse scenarios. This study provides a conceptual foundation for the technical and ethical development of MAS in the era of AI.

KEY-WORDS: Artificial intelligence; intelligent systems automation; multi-agent system architecture; Model Context Protocol – MCP

Revista Processando o Saber

eISSN 2179-5150 · Vol 18, n. 01, 2026
Multidisciplinar · DOI · Revisão por pares

Faculdade de Tecnologia Praia Grande – FATEC

Períodicidade: Anual
revista@fatecpg.edu.br

Recebido: Jan 2026

Aceito: Mar 2026

Publicado: Jun 2026

URL: <https://www.fatecpg.edu.br/revista/index.php/ps/article/view/422>

DOI: <https://doi.org/10.5281/zenodo.19930775>.



RESUMO

O avanço recente da Inteligência Artificial (IA), especialmente dos modelos generativos, ampliou as aplicações de sistemas inteligentes em ambientes distribuídos. Nesse cenário, os Multi-Agent Systems (MAS) emergem como uma estratégia-chave para coordenar agentes autônomos em tarefas complexas. O presente artigo tem como objetivo principal apresentar o estado da arte dos MAS sob a ótica da Inteligência Artificial contemporânea, com especial ênfase na integração com LLMs e no papel emergente do protocolo MCP como elemento facilitador dessa articulação. A metodologia adotada é de caráter qualitativo, ancorada em revisão narrativa da literatura, com levantamento de publicações em bases como Google Scholar e Scopus. Complementarmente, usou-se Excel para segmentação e análise crítica dos documentos. As conclusões desta investigação indicam que os MAS vêm se sofisticando frente à crescente complexidade dos ambientes distribuídos, especialmente com apoio de IA generativa. A integração com LLMs e o surgimento do MCP como protocolo de comunicação têm papel central nesse avanço. O MCP viabiliza interoperabilidade e decisões autônomas distribuídas, sendo promissor apesar de sua adoção inicial. Ainda há desafios, e pesquisas futuras devem avaliar sua eficácia em cenários mais diversos. Este estudo fornece uma base conceitual para o desenvolvimento técnico e ético dos MAS na era da IA.

INTRODUÇÃO

Os *Multi-Agent Systems* (MAS) existem há várias décadas e se tornaram mais inteligentes com o tempo. No início dos anos 2000, quando Sistemas Multiagentes eram usados principalmente em testes e simulações simples, pesquisadores se concentraram na formalização das bases teóricas e na padronização dos protocolos de comunicação entre agentes. Nesse contexto, destacou-se a consolidação da FIPA-ACL (*Foundation for Intelligent Physical Agents – Agent Communication Language*) como referência para a troca de mensagens estruturadas e semânticas entre entidades autônomas em ambientes distribuídos (Facchini, 2022).

A partir de 2010, os Sistemas Multiagentes passaram a ser implementados de forma mais intensiva em cenários de aplicação reais, especialmente em domínios complexos e dinâmicos como a gestão de tráfego urbano e sistemas de apoio à decisão na saúde (Wu et al., 2025). Essa fase representou a passagem dos MASs de ambientes simulados para aplicações reais, demandando maior robustez diante da incerteza.

Em 2015, o advento de tecnologias como *Distributed Ledger Technologies* (DLTs), com destaque para o *blockchain*, introduziu novas possibilidades para a descentralização e a confiabilidade das interações entre agentes, possibilitando a criação de ecossistemas multiagentes baseados em contratos inteligentes (*smart contracts*), com ênfase na transparência, rastreabilidade e imutabilidade das ações no sistema (Facchini, 2022). Esses avanços favoreceram o surgimento de estruturas autônomas como as DAOs (*Decentralized Autonomous Organizations*), apoiadas por arquiteturas distribuídas que permitem maior escalabilidade e resiliência, impulsionando novos modelos de governança e colaboração entre agentes.

Já entre 2024 e 2025, observou-se uma transformação significativa nos sistemas inteligentes impulsionada pelos avanços da Inteligência Artificial (IA) generativa, especialmente no contexto dos *Large Language Models* (LLMs). Esses modelos têm demonstrado capacidades notáveis de raciocínio, planejamento e adaptação, aproximando-se progressivamente de comportamentos cognitivos humanos. Ao tornar-se possível que agentes sejam capazes de inferir intenções, crenças e objetivos de outros agentes — humanos ou artificiais — os LLMs viabilizam a execução de tarefas complexas em ambientes dinâmicos, contribuindo para o surgimento de agentes autônomos mais flexíveis e menos dependentes de regras estáticas (Du et al., 2024).

Diferentemente dos agentes tradicionais, tais agentes são aptos a operar em cenários de alta incerteza, promovendo uma reconfiguração dos paradigmas de atuação em sistemas

multiagentes. Nesse contexto, destaca-se o surgimento e a valorização das arquiteturas multiagentes como resposta às novas demandas computacionais por soluções distribuídas, escaláveis e adaptativas. Tais arquiteturas possibilitam a coordenação de múltiplos agentes autônomos em ambientes heterogêneos e dinâmicos, otimizando a resolução de problemas complexos em diversas áreas do conhecimento (Du et al., 2024).

Os *Multi-Agent Systems* (MAS) são constituídos por coleções de agentes autônomos que interagem entre si com um ambiente compartilhado, visando atingir metas individuais e coletivas (Liu, 2024). Fundamentam-se em princípios-chave que lhes conferem propriedades singulares: 1) a autonomia, que permite aos agentes tomar decisões de forma independente com base em suas percepções e objetivos; 2) a cooperação, essencial para alcançar metas comuns por meio de comunicação, coordenação, consenso e informação (Wu et al., 2025); 3) a adaptabilidade, que os capacita a ajustar estratégias conforme as condições ambientais e as exigências das tarefas, especialmente em cenários dinâmicos com reconhecimento de contexto (Liu, 2024); 4) a escalabilidade, que viabiliza a distribuição de tarefas complexas entre múltiplos agentes, superando as limitações operacionais de sistemas baseados em agentes únicos (Du et al., 2024).

A partir destes conceitos fundamentais apresentados, se estrutura a base para a análise dos sistemas multiagentes sob a ótica da inteligência artificial generativa. O presente artigo tem como objetivo principal apresentar o estado da arte dos MAS sob a ótica da Inteligência Artificial contemporânea, com especial ênfase na integração com LLMs e no papel emergente do protocolo MCP como elemento facilitador dessa articulação. A tese central deste trabalho é que os sistemas multiagentes baseados em IA representam uma evolução significativa nas formas de coordenação e execução de tarefas complexas em ambientes computacionais dinâmicos.

O problema abordado neste estudo é a carência de estudos atualizados, em língua portuguesa, que sistematizem os avanços recentes do MCP como protocolo padrão de comunicação entre múltiplos agentes. Diante disso, realizou-se uma revisão de estudos internacionais de alta relevância que pudessem não só apresentar uma compreensão geral da trajetória mais recente do MAS enquanto tecnologia, mas também identificar possíveis tendências dessa tecnologia para o futuro. Portanto, a primeira parte deste artigo examina os fundamentos conceituais que caracterizam os MAS, além de apresentar um panorama geral da evolução destes sistemas entre 2000 e 2025. Em seguida, descreve-se a metodologia utilizada e finalmente, nos resultados, o artigo se aprofunda na análise do MCP, procurando explicar o contexto em que esse *framework* se insere no cenário atual dos MAS.

1. FUNDAMENTAÇÃO TEÓRICA

O conceito central de um MAS baseia-se na noção de agente autônomo (Du et al., 2024), definido como uma entidade computacional que percebe seu ambiente por meio de sensores e age sobre ele por meio de atuadores, de forma independente (Adams, 2001). Essa definição vai além da simples execução de comandos pré-programados, pois o agente autônomo é capaz de tomar decisões baseadas em regras, heurísticas ou processos de aprendizado, considerando uma leitura em tempo real do ambiente (Figura 1).

Figura 1 - Estrutura de interconexão entre agentes e seu meio



Fonte: do autor, gerada com Inteligência Artificial (ChatGPT 4o - OpenAI).

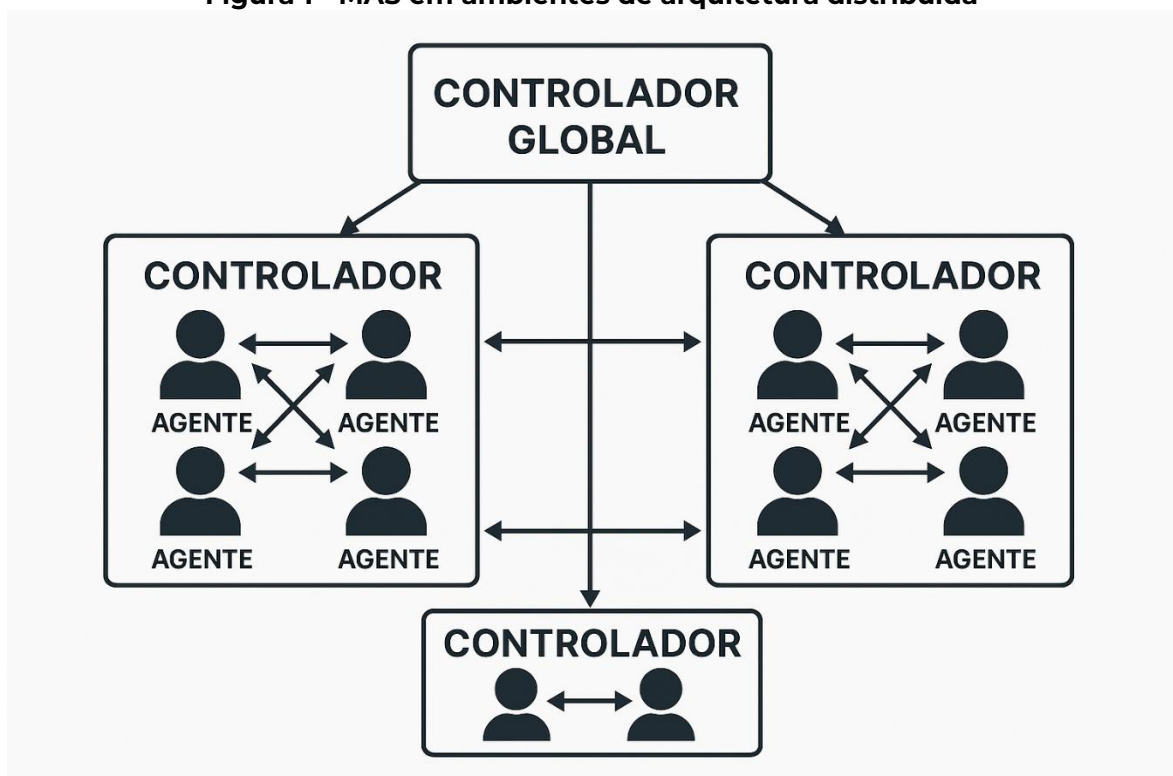
Como destacado por Jin et al. (2022), a robustez desses agentes em sistemas abertos depende diretamente da capacidade de adaptação a mudanças dinâmicas e da manutenção de colaboração eficiente mesmo sob condições incertas e adversas. Em MAS, tais agentes frequentemente operam sob incerteza e tempo limitado, exigindo capacidade de raciocínio deliberativo ou reativo para responder adequadamente às mudanças nas condições ambientais (Du et al. 2024).

Os Sistemas Multiagentes (MAS) representam uma abordagem computacional fundamentada na composição de agentes autônomos que operam em um ambiente compartilhado. Esses agentes interagem entre si, visando alcançar tanto objetivos individuais

quanto metas coletivas, conforme destacado por Du et al. (2024). A estrutura dos MAS permite representar sistemas nos quais múltiplos agentes, com conhecimentos e funções distintas, atuam de maneira coordenada para realizar tarefas que envolvem múltiplos elementos espaciais, temporais e decisórios, sendo particularmente eficaz em domínios onde a descentralização e a tomada de decisão local são desejáveis, como logística inteligente, gerenciamento de tráfego urbano, sistemas energéticos distribuídos, coordenação de robôs em ambientes industriais e exploração espacial autônoma (Liu, 2024).

Essa autonomia computacional é o que diferencia os MAS de sistemas convencionais centralizados, pois garante que os agentes possam operar mesmo quando desconectados de um controlador global, o que é fundamental para aplicações em ambientes distribuídos, como veículos autônomos (Jin et al., 2023), *drones* cooperativos e redes de sensores inteligentes (Shamshirband et al., 2013).

Figura 1 - MAS em ambientes de arquitetura distribuída



Fonte: Imagem gerada pelo autor com uso de Inteligência Artificial (*ChatGPT 4o - OpenAI*)

Além disso, no plano da arquitetura computacional, os MAS podem integrar Large Language Models (LLMs) como núcleos de raciocínio. LLMs são modelos de inteligência artificial treinados com grandes volumes de texto, capazes de compreender e gerar linguagem natural de forma sofisticada. Eles utilizam essas capacidades para interpretar linguagem natural,

gerar respostas contextualmente relevantes e formular estratégias de ação. Como destacado por Wu et al. (2025), a integração de LLMs em sistemas multiagentes tem o potencial de aprimorar significativamente a coordenação, a comunicação e o planejamento coletivo entre agentes autônomos. No backend, esses modelos funcionam como mecanismos deliberativos, capazes de analisar instruções complexas, simular cenários, inferir consequências e propor sequências de ações otimizadas de maneira autônoma e adaptativa. Essa autonomia permite que cada agente tome decisões com base em sua percepção do ambiente e em seus próprios objetivos, sem depender de controle centralizado (Du et al., 2024). Tal característica confere aos MAS uma robustez importante frente a falhas locais e mudanças inesperadas.

Além da autonomia, a cooperação é outro pilar essencial dos MAS (Liu, 2024). Os agentes são capazes de coordenar ações, negociar, compartilhar informações e alcançar consensos com outros agentes para atingir objetivos comuns (Liu, 2024). Esse comportamento colaborativo é habilitado por protocolos de comunicação e estratégias de coordenação que permitem sincronizar atividades e distribuir eficientemente as tarefas. Entre os protocolos mais utilizados destacam-se o FIPA-ACL (*Foundation for Intelligent Physical Agents – Agent Communication Language*), que define padrões para a troca de mensagens entre agentes, e o KQML (*Knowledge Query and Manipulation Language*), voltado para a troca de informações baseadas em conhecimento (Facchini, 2022).

Esses protocolos especificam estruturas semânticas e sintáticas para mensagens, bem como atos de fala como 'informar', 'pedir', 'confirmar' e 'propor', possibilitando que agentes heterogêneos negociem, cooperem e coordenem ações mesmo em ambientes distribuídos e com diferentes arquiteturas internas (Liu, 2024). Além disso, protocolos de coordenação, como o *Contract Net Protocol*, e mecanismos de leilão distribuído são amplamente empregados para a alocação dinâmica de tarefas, contribuindo para a eficiência coletiva do sistema (Adams, 2001).

A adaptabilidade dos agentes em um MAS é um diferencial crucial em ambientes dinâmicos, onde as condições e as exigências das tarefas podem mudar continuamente (Du et al., 2024). Os agentes adaptam suas estratégias de atuação com base no reconhecimento de contexto e aprendizado contínuo, o que é vital para a eficácia do sistema a longo prazo, pois permite que o MAS se mantenha funcional e com alta acurácia mesmo diante de variações imprevisíveis no ambiente (Du et al., 2024). Essa capacidade adaptativa é essencial para lidar com tarefas não estacionárias, novas condições operacionais e requisitos emergentes, reduzindo a dependência de reprogramações manuais e aumentando a resiliência do sistema frente a falhas, ambiguidade e mudanças de objetivos ao longo do tempo (Berducci et al., 2024).

Por fim, a escalabilidade dos MAS permite que tarefas complexas sejam decompostas e atribuídas a múltiplos agentes, ampliando o poder computacional coletivo do sistema e superando as limitações de arquiteturas centralizadas ou de agente único (Du et al., 2024). Isso se traduz na capacidade dos MAS de aumentar progressivamente o número de agentes e a diversidade de funções sem comprometer a eficiência global. À medida que novos agentes são adicionados, o sistema pode redistribuir tarefas, otimizar recursos e manter a qualidade das decisões de forma descentralizada. Essa característica é fundamental em aplicações de grande escala, como redes de veículos autônomos, sistemas de logística global e redes energéticas inteligentes, em que a necessidade de expansão dinâmica e adaptação rápida é crítica para garantir desempenho e resiliência (Wu et al., 2024).

Em síntese, a análise dos fundamentos conceituais dos MAS revela que a autonomia, a cooperação, a adaptabilidade e a escalabilidade não são apenas propriedades teóricas, mas requisitos operacionais concretos que determinam a viabilidade desses sistemas em ambientes reais. Na percepção deste autor, o desafio mais urgente reside na tensão entre autonomia e confiabilidade: à medida que os agentes se tornam mais independentes — especialmente com o suporte de LLMs —, cresce proporcionalmente a necessidade de mecanismos de verificação e controle que garantam decisões seguras e auditáveis. A cooperação mediada por protocolos como FIPA-ACL, embora consolidada, mostra-se insuficiente para as demandas semânticas dos agentes baseados em IA generativa, sugerindo que a próxima fronteira dos MAS dependerá de protocolos mais expressivos e interoperáveis.

1.1 HISTÓRICO DOS *MULTI-AGENT SYSTEMS*

A evolução dos Sistemas Multiagentes (MAS) ao longo das últimas duas décadas revela um percurso de sofisticação progressiva, marcado inicialmente por um amadurecimento teórico e, posteriormente, por uma integração crescente com técnicas de inteligência computacional. No início dos anos 2000, os MAS se consolidaram como uma abordagem promissora para a modelagem de sistemas distribuídos compostos por agentes autônomos capazes de interagir, cooperar ou competir entre si (Adams, 2001). As primeiras aplicações práticas destes sistemas se concentraram em ambientes controlados, apoiadas por protocolos padronizados como o FIPA-ACL, que normatizou a comunicação entre agentes. Todavia, a partir de 2010, intensificou-se o interesse na superação de limitações desses sistemas em contextos dinâmicos

e incertos, impulsionando a incorporação de técnicas de Inteligência Computacional (IC) nos MAS (Penserini et al., 2006).

Esse movimento resultou na consolidação do campo conhecido como MCI (*Multi-Agent System-based Computational Intelligence*), cuja proposta era aprimorar a capacidade adaptativa e preditiva dos MAS por meio da integração com métodos como lógica *fuzzy*, redes neurais e aprendizado por reforço. Tal abordagem demonstrou notável eficácia em aplicações voltadas à segurança cibernética, especialmente na detecção de intrusões em redes sem fio (*Wireless Intrusion Detection and Prevention Systems – WIDPS*). Conforme demonstrado por Shamshirband et al. (2013), as arquiteturas colaborativas baseadas em MCI — ou Co-WIDPS — apresentaram desempenho superior na identificação de padrões maliciosos, ao reduzirem taxas de falsos positivos e ampliarem a precisão dos diagnósticos em ambientes distribuídos.

A década de 2010 testemunhou, ainda, avanços significativos na aplicação prática dos MCI em ambientes reais. Um exemplo notável é o sistema *Sensor Network-based Counter-Sniper* (C-Sniper), adotado pelo Exército dos Estados Unidos, que exemplifica o uso bem-sucedido de agentes colaborativos e técnicas de IC para detectar e localizar ameaças em tempo real em zonas urbanas (Shamshirband et al., 2013). Em consonância com essa evolução, observa-se também a transição dos MAS para aplicações de larga escala em áreas como a robótica de consumo e os veículos autônomos.

Essa mudança refletiu a crescente centralidade da tomada de decisão autônoma em ambientes dinâmicos e incertos. Nesse contexto, os agentes autônomos precisaram superar os comportamentos rigidamente programados para atuarem de forma eficaz em ambientes descentralizados e distribuídos, passando a incorporar habilidades de tomada de decisão em tempo real, que possibilitam sua adaptação dinâmica diante de mudanças nas condições e de informações incompletas (Moharir et al., 2019).

Dando continuidade a essa trajetória de aprimoramento dos MAS, os anos 2020 inauguraram uma nova etapa com a integração da Inteligência Artificial Generativa (GenAI), especialmente por meio dos LLMs, que ampliaram significativamente as capacidades desses sistemas em contextos colaborativos com agentes humanos (Liu et al., 2023). Inicialmente, a IA modelava agentes individuais, considerando outros agentes como parte do ambiente (Gal; Grozs, 2022). A transição para grupos mistos exigiu novas abordagens capazes de representar os estados mentais, intenções e as capacidades humanas de decisão e comunicação. Com isso, os LLMs passaram a ser explorados de forma mais concreta em aplicações práticas de MAS.

Um dos principais campos dessa aplicação tem sido a direção autônoma distribuída, otimizando a coordenação entre veículos, infraestrutura inteligente e interfaces virtuais,

incorporando seres humanos como agentes virtuais no ecossistema (Wu et al., 2025). Ao operar sob a lógica dos MAS Conscientes do Contexto (CA-MAS), esses sistemas seguem ciclos funcionais estruturados em fases como "Sentir, Aprender, Raciocinar, Prever e Agir", sendo os LLMs essenciais para as etapas de aprendizado situacional e raciocínio adaptativo (Liu et al., 2023). Em implementações práticas, os LLMs têm assumido o papel de módulos deliberativos capazes de traduzir instruções linguísticas em rotas planejadas e decisões estratégicas em tempo real, o que amplia significativamente a autonomia e a capacidade de adaptação dos agentes (Gal; Grosz, 2022).

Apesar das vantagens, a incorporação de LLMs em MAS também traz desafios, como o fenômeno das alucinações, no qual LLMs geram inferências factualmente incorretas com alta fluência linguística, comprometendo a integridade da coordenação em ambientes multiagente. Esse comportamento, particularmente crítico em contextos distribuídos, demanda mecanismos formais de verificação semântica e controle de confiança interagente (Islam; Lauscher; Glavas, 2025). Frente a isso, estudos recentes têm priorizado abordagens de segurança adaptativa, reformulando o problema sob o arcabouço dos CPOMDPs (*Constrained Partially Observable Markov Decision Processes*) e incorporando Funções de Barreira de Controle (CBF) como salvaguardas formais para garantia de segurança mesmo sob incerteza e parcialidade observacional (Lindemann; Dimarogonas, 2019). Complementarmente, o avanço do raciocínio latente e arquiteturas como o paradigma *Coconut* (*Chain of Continuous Thought*) propõem uma expansão do raciocínio simbólico, permitindo que LLMs operem iterativamente em espaços latentes contínuos, o que abre novas possibilidades para escalabilidade cognitiva e raciocínio composicional em tempo de inferência (Geiping et al., 2025).

É neste sentido que abordagens como DAOs (Organizações Autônomas Descentralizadas) têm sido exploradas em conjunto com MAS para promover estruturas organizacionais descentralizadas, autoexecutáveis e resistentes a falhas em ambientes distribuídos. Enquanto os MAS fornecem uma base para agentes autônomos colaborarem em tarefas complexas, os DAOs introduzem mecanismos de governança baseados em contratos inteligentes que permitem a tomada de decisões automatizada e transparente. Essa integração é particularmente relevante para aplicações em IA e análise de dados em ambientes de múltiplas partes interessadas, quando a confiança, a auditabilidade e a descentralização são essenciais (Berducci et al., 2024).

Do ponto de vista analítico, a trajetória histórica dos MAS demonstra que cada salto tecnológico, da padronização FIPA-ACL à incorporação de LLMs, foi motivado por limitações concretas do paradigma anterior. Já a integração com IA generativa representa a transformação

mais substancial dessa trajetória, pois não se trata apenas de um aprimoramento incremental, mas de uma mudança qualitativa na capacidade cognitiva dos agentes. No entanto, essa sofisticação traz riscos proporcionais: o fenômeno das alucinações em LLMs e a dificuldade de garantir segurança em ambientes críticos são, na visão deste estudo, os obstáculos mais desafiadores a serem superados. Tecnologias como DAOs e mecanismos de raciocínio latente parecem sustentáveis a médio prazo, mas sua consolidação dependerá de avanços em verificação formal e governança descentralizada.

2. PROCEDIMENTOS METODOLÓGICOS

A presente pesquisa adota uma abordagem qualitativa, de natureza exploratória, com delineamento fundamentado na revisão narrativa da literatura. Esse tipo de revisão busca construir uma análise crítica e interpretativa sobre um determinado campo de conhecimento, sem o uso de protocolos sistemáticos rígidos, permitindo ao pesquisador articular diferentes fontes teóricas e identificar tendências emergentes.

A revisão narrativa foi estruturada com base na seleção de publicações científicas e técnicas relacionadas aos MASs no contexto da Inteligência Artificial, buscando estudos publicados entre os anos de 2000 e 2025. As buscas foram realizadas em bases de dados acadêmicas amplamente reconhecidas, como IEEE Xplore, Google Scholar e Scopus, além de documentos técnicos disponibilizados por comunidades científicas e empresas. Os termos utilizados para a recuperação dos estudos incluíram: "*multi-agent systems*", "*artificial intelligence*", "*agent-based architecture*", "*LLMs*", "*LLM orchestration in MAS*", "*agent coordination*". Foram considerados elegíveis apenas os artigos redigidos em português ou inglês, disponíveis na íntegra e com aderência temática aos objetivos da pesquisa. Estudos desatualizados ou que não apresentassem conexão direta com a problemática investigada foram excluídos da análise.

Para a seção específica dedicada ao Model Context Protocol (MCP), foram selecionados dois documentos centrais: Szeider (2025) e Anthropic (2024). A escolha desses materiais justificou-se por sua atualidade e por apresentarem sínteses concisas, porém aprofundadas, sobre o MCP e sua integração com LLMs — complementando as demais fontes utilizadas ao longo do artigo para a fundamentação teórica e histórica dos MAS. O documento da Anthropic, em particular, foi incluído por se tratar da organização responsável pelo lançamento do MCP,

descrito como “um novo padrão para conectar assistentes de IA a sistemas” (**tradução nossa**) (Anthropic, 2024, s.p.).

Na etapa subsequente, os textos selecionados foram sistematicamente organizados em planilhas Excel, nas quais se aplicaram estratégias de leitura ativa, incluindo a elaboração de linhas do tempo conceituais, mapas mentais e fichamentos analíticos. O objetivo desse procedimento foi permitir a articulação interpretativa e integrativa dos dados ao longo do desenvolvimento do estudo. Por fim, o ChatGPT-4o foi empregado para a criação de imagens ilustrativas destinadas a exemplificar visualmente determinados conceitos centrais da análise.

3. RESULTADOS E DISCUSSÃO - PERSPECTIVAS E TENDÊNCIAS ATUAIS DOS MULTI-AGENT SYSTEMS

Como visto na seção anterior, a implementação de LLMs em Sistemas Multiagente surgiu como uma resposta estratégica à crescente necessidade de dotar agentes autônomos de capacidades avançadas de linguagem natural e raciocínio contextual. Antes da consolidação de protocolos padronizados de comunicação entre agentes, embora os LLMs demonstrassem desempenho notável em tarefas linguísticas, ainda apresentavam limitações severas em operações de raciocínio lógico formal. Essa lacuna tornou imperativa a busca por soluções capazes de complementar os LLMs com mecanismos externos de raciocínio simbólico, uma necessidade especialmente sentida em ambientes de MAS, onde decisões precisas e interações complexas são requisitos fundamentais (Szeider, 2025).

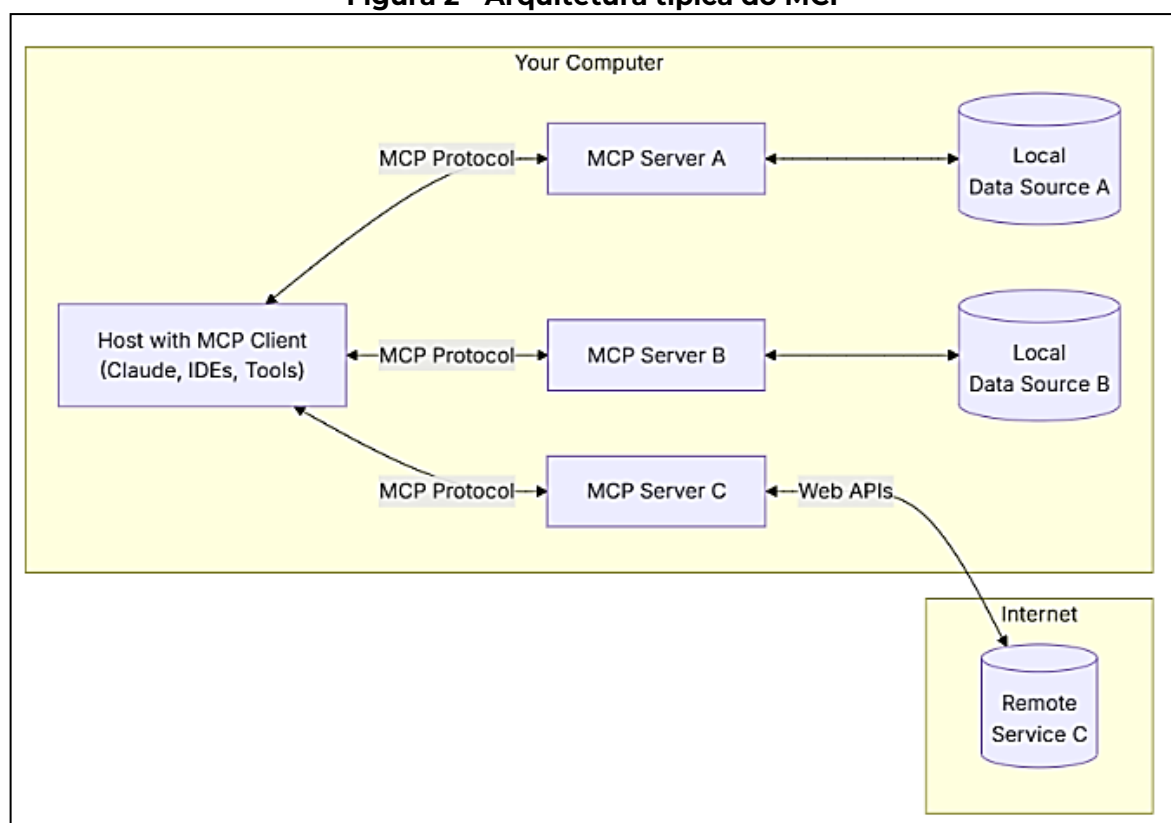
Nesse contexto, surgiram diversos protótipos com o objetivo de conectar LLMs a solucionadores simbólicos para mitigar suas fragilidades. Entre essas abordagens, destacam-se o PRoC3S para planejamento robótico, sistemas baseados em síntese de programas assistidos por contraexemplos, e projetos como o SATLM e o LOGIC-LM focados na tradução de linguagem natural para formatos de solucionadores. O *framework* Lemur também emergiu como uma tentativa de estruturar a síntese de programas de maneira LLM-agnóstica. Embora pioneiras, essas iniciativas apresentavam restrições consideráveis em termos de flexibilidade, integração e interoperabilidade, limitando sua aplicabilidade a domínios específicos (Szeider, 2025).

As principais limitações das abordagens iniciais estavam relacionadas à rigidez dos *pipelines* fixos, à integração restrita para casos de uso isolados, à falta de suporte para interações iterativas generalizadas e à inexistência de um padrão aberto que possibilitasse a

interoperabilidade entre diferentes agentes e solucionadores. Essa ausência de padronização dificultava a criação de ecossistemas colaborativos robustos e sustentáveis, elementos essenciais em implementações de MAS de larga escala que demandam coordenação sofisticada entre agentes heterogêneos.

A introdução do MCP, lançado pela *Anthropic* em novembro de 2024, representou um marco disruptivo nesse cenário. O MCP estabeleceu um padrão aberto e universal para conectar LLMs a sistemas externos, proporcionando uma arquitetura flexível baseada em comunicação cliente-servidor com estado. Essa estrutura permite que dados e capacidades computacionais sejam expostos via servidores padronizados, aos quais os agentes ou aplicações de IA se conectam como clientes, permitindo trocas de informações estruturadas e interativas, algo inédito nos protocolos anteriores. O protocolo também possui integração nativa com o *Claude Desktop*, uma aplicação desenvolvida pela própria *Anthropic* que atua como uma interface local para o uso de modelos *Claude* no ambiente de trabalho. O *Claude Desktop* permite que usuários interajam com LLMs diretamente no computador, com suporte a plugins e conexões seguras a servidores MCP locais, facilitando o acesso a arquivos, sistemas e bancos de dados da própria máquina. A figura abaixo ilustra o funcionamento do MCP em sistemas de multiagentes:

Figura 2 - Arquitetura típica do MCP



Fonte: *Anthropic*, 2024.

O MCP organiza-se em três componentes essenciais: (i) a especificação técnica, que detalha os formatos, métodos e transportes suportados (como JSON-RPC 2.0, JSON Schema, stdio e HTTP/SSE); (ii) os SDKs (*Software Development Kits*) oficiais, disponíveis em diversas linguagens (*TypeScript, Python, Java, Kotlin, C#, Rust e Swift*), que simplificam a criação de clientes e servidores locais ao abstrair os detalhes de transporte e de *JSON-RPC* e; (iii) o repositório *open-source* de servidores, que oferece implementações de referência e links para integrações de terceiros, mostrando como expor dados via MCP. Essa estrutura modular permite que qualquer organização conecte rapidamente seus *datasets* mais relevantes a ferramentas de IA utilizando integrações prontas para serviços populares como *Google Drive, Slack, GitHub e Postgres*, e acelera a adoção do protocolo graças às implementações de servidores MCP já disponíveis. A citação abaixo procura fazer uma analogia acerca do funcionamento do MCP:

MCP is an open protocol that standardizes how applications provide context to LLMs. Think of MCP like a USB-C port for AI applications. Just as USB-C provides a standardized way to connect your devices to various peripherals and accessories, MCP provides a standardized way to connect AI models to different data sources and tools (Anthropic, 2024, s.p.).

A sua arquitetura do MCP trouxe avanços fundamentais ao suportar interações iterativas entre agentes e solucionadores, viabilizando adaptações dinâmicas conforme a evolução das tarefas. Empresas como *Block* e *Apollo* já integraram o MCP aos seus sistemas, enquanto plataformas de desenvolvimento como *Zed, Replit, Codeium e Sourcegraph* trabalham para incorporar o protocolo, evidenciando sua importância estratégica. Como ressaltado por Dhanji R. Prasanna, CTO da Block, “*Open technologies like the Model Context Protocol are the bridges that connect AI to real-world applications, ensuring innovation is accessible, transparent, and rooted in collaboration*” (Anthropic, 2024, s.p.). Essa adoção expressiva evidencia a capacidade do MCP de atender às exigências operacionais de sistemas reais, promovendo soluções interoperáveis, escaláveis e orientadas para a colaboração multiagente (Szeider, 2025).

É sob estas condições que o MCP emerge como um *framework* essencial para a próxima geração de MAS, pois sua padronização e flexibilidade não apenas superam as restrições das abordagens anteriores, como também oferecem a infraestrutura necessária para que agentes equipados com LLMs possam interagir de maneira mais eficaz, segura e inteligente com o ecossistema computacional ao seu redor. A Figura 3 ilustra de maneira esquemática a arquitetura típica de funcionamento do MCP, destacando a relação entre clientes, servidores e fontes de dados locais ou remotas.

Assim, o MCP sinaliza uma direção promissora para o futuro da integração entre LLMs e ecossistemas computacionais distribuídos. À medida que modelos de linguagem se tornam mais capazes de executar raciocínios complexos, agir de forma autônoma e interagir em ambientes abertos, a necessidade de protocolos estáveis, interoperáveis e auditáveis deve se intensificar. O MCP, ao oferecer uma camada de padronização e abstração, poderá tornar-se o elo entre agentes cada vez mais sofisticados e infraestruturas de dados heterogêneas, não apenas em domínios corporativos, mas também em contextos acadêmicos, governamentais e industriais de grande escala.

Entretanto, é importante reconhecer, como outra possível tendência, que o avanço exponencial dos LLMs pode, por exemplo, levar ao surgimento de arquiteturas mais dinâmicas ou autossuficientes, que eventualmente dispensem intermediários como o MCP em determinados contextos. Ainda assim, enquanto os sistemas permanecerem dependentes de dados externos não estruturados ou contextos altamente específicos, protocolos como o MCP deverão manter sua relevância. A tendência, portanto, é que esse tipo de infraestrutura evolua para suportar níveis mais altos de segurança, versionamento de contexto, interoperabilidade semântica e adaptação a arquiteturas distribuídas com múltiplos agentes autônomos operando de forma coordenada.

Em termos críticos, os resultados desta investigação indicam que o MCP representa, na avaliação deste autor, a tecnologia de integração mais promissora e sustentável a médio prazo para os MAS baseados em LLMs. Sua arquitetura aberta e modular, já adotada por empresas de grande porte, confere ao MCP vantagem significativa em relação às abordagens anteriores, que eram fragmentadas e proprietárias. Contudo, é preciso reconhecer que a dependência de um único protocolo, ainda em estágio inicial de maturação, introduz riscos de concentração tecnológica. O desafio mais urgente, portanto, não é apenas técnico, mas também estratégico: garantir que a evolução do MCP ocorra de forma colaborativa e aberta, evitando a captura por interesses corporativos específicos, e que mecanismos robustos de segurança e auditabilidade acompanhem a crescente autonomia dos agentes em ambientes distribuídos.

4. CONSIDERAÇÕES FINAIS

O objetivo principal desta investigação foi apresentar o estado da arte dos MAS sob a ótica da Inteligência Artificial contemporânea, com especial ênfase na integração com LLMs e no papel emergente do protocolo MCP como elemento facilitador dessa articulação. É possível

afirmar que os resultados obtidos ao longo da revisão narrativa indicam uma trajetória clara de sofisticação técnica e de crescente integração com tecnologias de IA generativa. Nesse contexto, os sistemas multiagentes têm incorporado mecanismos de raciocínio, adaptação e coordenação mais complexos, e o MCP, ao propor um protocolo unificado de comunicação entre agentes e sistemas externos, destaca-se como solução promissora para viabilizar fluxos distribuídos de decisão autônoma.

Tomados em conjunto, estes resultados sugerem que a evolução dos MAS está intimamente ligada à capacidade de articulação entre componentes especializados e à possibilidade de adaptação contínua ao contexto. Nesse cenário, o MCP, funciona como elo técnico entre agentes e fontes heterogêneas de dados, promovendo interoperabilidade e escalabilidade — duas exigências fundamentais para que tais arquiteturas se consolidem em ambientes produtivos reais. Ainda que a consolidação do MCP esteja em estágio inicial, sua adoção por grandes plataformas e comunidades de desenvolvimento sinaliza um caminho sólido para o estabelecimento de um ecossistema de agentes inteligentes mais conectado, flexível e colaborativo.

Apesar dos avanços e das perspectivas promissoras, muitas perguntas permanecem em aberto. Como garantir a confiabilidade das decisões geradas por LLMs em ambientes críticos? Quais os limites da autonomia dos agentes em contextos com conflitos éticos ou normativos? Como assegurar a interoperabilidade sem comprometer a segurança dos dados? O MCP será capaz de sustentar o crescimento de ecossistemas de larga escala com múltiplos agentes heterogêneos? E até que ponto os agentes poderão adaptar-se a ambientes totalmente novos sem intervenção humana? Essas questões não apenas evidenciam as lacunas teóricas e técnicas ainda existentes, como também orientam os rumos para investigações futuras.

Novas pesquisas devem ser realizadas para investigar a efetividade do MCP em cenários com maior variabilidade de agentes, uso de fontes de dados não convencionais e presença de objetivos concorrentes. Nesse sentido, recomenda-se que futuros estudos adotem abordagens experimentais e comparativas, buscando avaliar o desempenho do MCP frente a outras arquiteturas de integração, bem como seu impacto em aspectos como segurança, latência e interpretabilidade.

REFERÊNCIAS

- ADAMS, J. A. Multiagent systems: a modern approach to distributed artificial intelligence. **AI Magazine**, v. 22, n. 2, p. 105-105, 2001. DOI: <https://doi.org/10.1609/aimag.v22i2.1567>.
- ANTHROPIC. Introducing the Model Context Protocol. [S. l.], 25 nov. 2024. Disponível em: <https://www.anthropic.com/news/model-context-protocol>. Acesso em: 25 abr. 2025.
- BERDUCCI, L. et al. Learning adaptive safety for multi-agent systems. In: **IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION (ICRA)**, 2024. [S. l.]: IEEE, 2024. p. 2859-2865. DOI: <https://doi.org/10.48550/arXiv.2309.10657>.
- DU, H. et al. A survey on context-aware multi-agent systems: techniques, challenges and future directions. **arXiv preprint**, arXiv:2402.01968, 2024. DOI: <https://doi.org/10.48550/arXiv.2402.01968>.
- FACCHINI, S. D. Decentralized autonomous organizations and multi-agent systems for artificial intelligence applications and data analysis. In: **PROCEEDINGS OF THE THIRTY-FIRST INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE (IJCAI-22)**. [S. l.], 2022. p. 5851-5852. DOI: <https://doi.org/10.24963/ijcai.2022/828>.
- GAL, K.; GROSZ, B. J. Multi-agent systems: technical & ethical challenges of functioning in a mixed group. **Daedalus**, v. 151, n. 2, p. 114-126, 2022. DOI: https://doi.org/10.1162/daed_a_01904.
- GEIPING, J. et al. Scaling up test-time compute with latent reasoning: a recurrent depth approach. **arXiv preprint**, arXiv:2502.05171, 2025. DOI: <https://doi.org/10.48550/arXiv.2502.05171>.
- ISLAM, S. O. ul; LAUSCHER, A.; GLAVAŠ, G. How much do LLMs hallucinate across languages? On multilingual estimation of LLM hallucination in the wild. **arXiv preprint**, arXiv:2502.12769, 2025. DOI: <https://doi.org/10.48550/arXiv.2502.12769>.
- JIN, D. et al. Tackling challenges of robustness measures for autonomous agent collaboration in open multi-agent systems. In: **PROCEEDINGS OF THE 55TH HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES**. [S. l.], 2022. p. 7585-7594. DOI: <https://doi.org/10.24251/HICSS.2022.911>.
- LINDEMANN, L.; DIMAROGONAS, D. V. Control barrier functions for multi-agent systems under conflicting local signal temporal logic tasks. **IEEE Control Systems Letters**, v. 3, n. 3, p. 757-762, 2019. DOI: <https://doi.org/10.1109/LCSYS.2019.2917975>.
- LIU, J. Multi-agent systems: studying coordination and cooperation mechanisms in multi-agent systems to achieve collective goals efficiently. **Journal of Artificial Intelligence Research**. [S. l.]: The Science Brigade Publishing Group, v. 4, n. 1, 2024. Disponível em: <https://thesciencebrigade.com/JAIR/article/view/98>. Acesso em: 15 mar. 2026.
- MOHARIR, M.; MAHALAKSHMI, A. S.; KUMAR, G. P. Autonomous agents: beginnings, innings and where we are headed. In: **Recent Advances in Computational Intelligence**. [S. l.], 2019. p. 255-261. DOI: https://doi.org/10.1007/978-3-030-12500-4_16.

PENSERINI, L. et al. From stakeholder intentions to software agent implementations. In: **Advanced Information Systems Engineering: 18th International Conference, CAiSE 2006, Luxembourg, Luxembourg, June 5-9, 2006, Proceedings. Lecture Notes in Computer Science**. Berlin: Springer, 2006. p. 465-479. DOI: <https://doi.org/10.1007/11767138>.

SHAMSHIRBAND, S. et al. An appraisal and design of a multi-agent system based cooperative wireless intrusion detection computational intelligence technique. **Engineering Applications of Artificial Intelligence**, v. 26, n. 9, p. 2105-2127, 2013. DOI: <https://doi.org/10.1016/j.engappai.2013.04.010>.

SZEIDER, S. MCP-solver: integrating language models with constraint programming systems. **arXiv preprint**, arXiv:2501.00539, 2025. DOI: <https://doi.org/10.48550/arXiv.2501.00539>.

WU, Y. et al. Multi-agent autonomous driving systems with large language models: a survey of recent advances. **arXiv preprint**, arXiv:2502.16804, 2025. DOI: <https://doi.org/10.48550/arXiv.2502.16804>.