Revista Ibero-Americana de Humanidades, Ciências e Educação — REASE

OPEN ACCESS

doi.org/10.51891/rease.v12i11.32036

# INTEGRAÇÃO ENTRE AGENTES INTELIGENTES CLÁSSICOS E AGENTES BASEADOS EM LLM: UMA REVISÃO SISTEMÁTICA DE ARQUITETURAS HÍBRIDAS

## BRIDGING CLASSICAL INTELLIGENT AGENTS AND LLM-BASED AGENTS: A SYSTEMATIC REVIEW OF HYBRID ARCHITECTURES

## INTEGRACIÓN ENTRE AGENTES INTELIGENTES CLÁSICOS Y AGENTES BASADOS EN LLM: UNA REVISIÓN SISTEMÁTICA DE ARQUITECTURAS HÍBRIDAS

**Gustavo Augusto Lima de Campos**[1]
**Thelmo Pontes de Araújo**[2]
**Thaynara Bezerra Lima de Almeida**[3]
**Caio Marques Silva**[4]
**João Lucas Simões Rocha da Silveira**[5]
**Alberto Luian de Lima Marques**[6]
**Anderson Herculano de Lima**[7]

RESUMO: Este artigo analisa e sintetiza diferentes formas de integração entre agentes inteligentes clássicos e agentes baseados em grandes modelos de linguagem (LLMs), por meio de uma revisão sistemática da literatura sobre arquiteturas híbridas. A metodologia foi fundamentada nas diretrizes de Kitchenham e Charters, utilizando o framework PICOC para orientar as questões de pesquisa e a diretriz PRISMA como referência para o relato do processo de seleção dos estudos. Foram analisados 20 trabalhos científicos, originalmente selecionados com foco em sistemas de recomendação e com a inclusão de estudos de outros domínios pertinentes à investigação arquitetural. A análise permitiu identificar três configurações principais: integração entre agentes de paradigmas distintos, hibridização de mecanismos clássicos e baseados em LLM no interior de agentes e combinações parciais de componentes computacionais. Os resultados evidenciaram diferentes níveis de integração, destacando-se aplicações em manufatura inteligente e limitações relacionadas à validação experimental, à confiabilidade e aos custos computacionais. No contexto complementar dos sistemas de recomendação, não foram identificadas, entre os estudos selecionados, arquiteturas que demonstrassem a integração direta entre agentes clássicos e agentes baseados em LLM independentes. Conclui-se que a caracterização dos papéis e mecanismos de interação entre agentes é fundamental para compreender as possibilidades e limitações dessas arquiteturas, permanecendo necessária a realização de avaliações comparativas mais abrangentes.

**Palavras-chave:** Agentes inteligentes. Grandes modelos de linguagem. Arquiteturas híbridas. Sistemas multiagentes. Revisão sistemática da literatura.



[1] Orientador. Doutor. Docente do curso de Ciência da Computação, Universidade Estadual do Ceará (UECE).
[2] Coorientador. Doutor. Docente do curso de Ciência da Computação, Universidade Estadual do Ceará (UECE).
[3] Aluna especial do curso de Pós-Graduação em Ciência da Computação, Universidade Estadual do Ceará (UECE). Graduada em Administração, Centro Universitário 7 de Setembro.
[4] Doutorando do Programa de Pós-Graduação em Engenharia de Teleinformática (PPGETI), Universidade Federal do Ceará (UFC). Docente, Universidade de Fortaleza (UNIFOR).
[5] Discente do curso de graduação em Ciência da Computação, Universidade Estadual do Ceará (UECE).
[6] Mestrando do Programa de Pós-Graduação em Ciência da Computação, Universidade Estadual do Ceará (UECE).
[7] Discente do curso de graduação em Ciência da Computação, Universidade de Fortaleza (UNIFOR).

ABSTRACT: This article analyzes and synthesizes different forms of integration between classical intelligent agents and large language model (LLM)-based agents through a systematic literature review of hybrid architectures. The methodology was based on Kitchenham and Charters' guidelines, using the PICOC framework to guide the research questions and the PRISMA guidelines as a reference for reporting the study selection process. Twenty scientific studies were analyzed, originally selected with a focus on recommender systems while also including studies from other domains relevant to the architectural investigation. The analysis identified three main configurations: integration between agents belonging to different paradigms, hybridization of classical and LLM-based mechanisms within individual agents, and partial combinations of computational components. The findings revealed different levels of integration, particularly in smart manufacturing applications, as well as limitations related to experimental validation, reliability, and computational costs. In the complementary context of recommender systems, none of the selected studies demonstrated direct integration between independent classical intelligent agents and LLM-based agents. The review concludes that characterizing agent roles and interaction mechanisms is essential to understanding the possibilities and limitations of these architectures, while more comprehensive comparative evaluations remain necessary.

**Keywords:** Intelligent agents. Large language models. Hybrid architectures. Multi-agent systems. Systematic literature review.

RESUMEN: Este artículo analiza y sintetiza diferentes formas de integración entre agentes inteligentes clásicos y agentes basados en grandes modelos de lenguaje (LLM), mediante una revisión sistemática de la literatura sobre arquitecturas híbridas. La metodología se fundamentó en las directrices de Kitchenham y Charters, utilizando el marco PICOC para orientar las preguntas de investigación y las directrices PRISMA como referencia para informar sobre el proceso de selección de los estudios. Se analizaron 20 trabajos científicos, seleccionados originalmente con énfasis en sistemas de recomendación e incluyendo estudios de otros dominios relevantes para la investigación arquitectónica. El análisis permitió identificar tres configuraciones principales: integración entre agentes de distintos paradigmas, hibridación de mecanismos clásicos y basados en LLM dentro de los agentes y combinaciones parciales de componentes computacionales. Los resultados evidenciaron diferentes niveles de integración, especialmente en aplicaciones de fabricación inteligente, así como limitaciones relacionadas con la validación experimental, la fiabilidad y los costes computacionales. En el contexto complementario de los sistemas de recomendación, no se identificaron, entre los estudios seleccionados, arquitecturas que demostraran una integración directa entre agentes inteligentes clásicos y agentes basados en LLM independientes. Se concluye que la caracterización de las funciones y los mecanismos de interacción entre agentes es fundamental para comprender las posibilidades y limitaciones de estas arquitecturas, y que siguen siendo necesarias evaluaciones comparativas más amplias.

**Palabras clave:** Agentes inteligentes. Grandes modelos de lenguaje. Arquitecturas híbridas. Sistemas multiagente. Revisión sistemática de la literatura.



## 1 INTRODUÇÃO

A evolução dos sistemas inteligentes tem sido impulsionada pela necessidade de processar grandes volumes de dados, interpretar informações provenientes de diferentes fontes e apoiar processos de tomada de decisão em contextos cada vez mais complexos. Tradicionalmente, esses sistemas empregam diferentes técnicas de Inteligência Artificial, incluindo abordagens baseadas em agentes inteligentes, capazes de perceber o ambiente,

processar informações e executar ações orientadas a determinados objetivos (RUSSELL; NORVIG, 2013; WANG et al., 2024).

Segundo Russell e Norvig (2013), um agente é uma entidade capaz de perceber seu ambiente por meio de sensores e agir sobre ele por intermédio de atuadores. Os autores também definem o agente racional como aquele que seleciona ações visando maximizar sua medida de desempenho esperada, considerando as informações disponíveis, seu conhecimento prévio e suas percepções do ambiente.

Agentes inteligentes clássicos podem empregar mecanismos de raciocínio simbólico, regras, planejamento e aprendizagem para estruturar processos de decisão e atuar em ambientes com objetivos e restrições previamente estabelecidos (RUSSELL; NORVIG, 2013). Entretanto, dependendo da abordagem adotada, esses agentes podem apresentar limitações diante de situações que exigem interpretação de informações não estruturadas, compreensão de intenções e tratamento de nuances semânticas da linguagem natural. Nesse cenário, conciliar informações subjetivas e contextuais com mecanismos estruturados de decisão constitui um desafio relevante para o desenvolvimento de sistemas inteligentes (WANG et al., 2024).

A ascensão dos grandes modelos de linguagem (LLMs) ampliou as possibilidades de aplicação da Inteligência Artificial, especialmente pela capacidade desses modelos de processar e gerar linguagem natural e interpretar informações expressas em diferentes contextos. Agentes baseados em LLM podem contribuir para a compreensão de solicitações, objetivos e preferências formulados de maneira não estruturada, ampliando as possibilidades de interação entre usuários e sistemas inteligentes (WANG et al., 2024; XI et al., 2025). Entretanto, sua utilização também apresenta desafios, incluindo a ocorrência de alucinações, custos computacionais, latência de inferência e dificuldades para garantir consistência e aderência a regras e restrições previamente estabelecidas (NISA et al., 2026; RAZA et al., 2026b). Essas limitações motivam a investigação de mecanismos capazes de articular as capacidades de processamento semântico dos LLMs com abordagens que ofereçam recursos estruturados de raciocínio, planejamento, controle e tomada de decisão.



Diversos problemas computacionais envolvem a necessidade de interpretar informações semânticas, frequentemente expressas em linguagem natural, e conciliá-las com regras, critérios e restrições associados à tomada de decisão. Esses problemas podem envolver preferências subjetivas, objetivos parcialmente especificados, requisitos contextuais e condições objetivas que precisam ser simultaneamente considerados na seleção de ações ou na elaboração de

soluções. Exemplos podem ser encontrados em sistemas de recomendação personalizada, planejamento de atividades, manufatura inteligente, logística e automação de ambientes (NAWARA; KASHEF, 2025; ZHAO, Z. et al., 2026; FARAHANI; KHAN; WUEST, 2026). Embora esses domínios apresentem características específicas, compartilham desafios relacionados à representação de informações heterogêneas, à interpretação de intenções e à realização de decisões consistentes com os objetivos e as restrições estabelecidas.

Nesse contexto, a integração entre agentes inteligentes clássicos e agentes baseados em LLM constitui uma perspectiva relevante para o desenvolvimento de arquiteturas híbridas. Essa integração pode envolver diferentes formas de organização arquitetural, desde a combinação de mecanismos clássicos e componentes baseados em LLM no interior de agentes até a coordenação entre agentes especializados de diferentes paradigmas, conforme ilustrado por abordagens presentes na literatura (WESNER et al., 2026; ZHAO, Z. et al., 2026; FARAHANI; KHAN; WUEST, 2026). A articulação dessas capacidades permite investigar aspectos como a distribuição de responsabilidades, a comunicação entre componentes, a coordenação de ações e o tratamento de restrições. Assim, o interesse científico pelas arquiteturas híbridas não se limita a um domínio específico, abrangendo diferentes contextos de aplicação e possibilidades de integração.



Diante desse cenário, torna-se relevante compreender como a literatura científica tem abordado a integração entre agentes inteligentes clássicos e agentes baseados em LLM, quais estratégias arquiteturais têm sido propostas, em quais domínios são aplicadas e quais vantagens, limitações e desafios são relatados. Para investigar essas questões, este artigo apresenta uma revisão sistemática da literatura fundamentada nas diretrizes de Kitchenham e Charters (2007), utilizando o PRISMA como referência para o relato do processo de seleção dos estudos (PAGE et al., 2021). A análise considera a amostra originalmente selecionada com foco em sistemas de recomendação, incluindo trabalhos de outros domínios pertinentes à investigação arquitetural.

Portanto, o objetivo deste estudo é analisar sistematicamente a literatura científica selecionada sobre a integração entre agentes inteligentes clássicos e agentes baseados em grandes modelos de linguagem, identificando os padrões arquiteturais empregados, os domínios de aplicação, as vantagens e limitações relatadas e os desafios associados às abordagens híbridas. De maneira complementar, busca-se examinar como esses paradigmas têm sido empregados em sistemas de recomendação, considerando as características e limitações das abordagens identificadas nesse domínio.

## 2 MÉTODOS

A metodologia utilizada neste trabalho de revisão sistemática da literatura foi baseada nas diretrizes propostas por Kitchenham e Charters (2007), utilizando a diretriz PRISMA como referência para o relato do processo de identificação, triagem e seleção dos estudos científicos. O planejamento original da pesquisa foi estruturado por meio do framework PICOC, com foco na integração entre agentes inteligentes clássicos e agentes baseados em grandes modelos de linguagem (LLMs) em sistemas de recomendação. A seleção também contemplou estudos de outros domínios que apresentassem características arquiteturais pertinentes à investigação. A partir da amostra selecionada, a análise foi ampliada para caracterizar diferentes estratégias de integração entre esses paradigmas, considerando seus padrões arquiteturais, domínios de aplicação, vantagens, limitações e desafios.

O planejamento original da revisão foi estruturado por meio do framework PICOC, que considera cinco elementos: população (Population), intervenção (Intervention), comparação (Comparison), resultados (Outcomes) e contexto (Context). A população (P) compreendeu sistemas de recomendação baseados em agentes inteligentes clássicos ou agentes baseados em LLM; a intervenção (I) correspondeu às arquiteturas híbridas que integram esses paradigmas; a comparação (C) considerou abordagens clássicas e abordagens baseadas em LLM; os resultados (O) contemplaram desempenho de recomendação, interpretabilidade, capacidade de raciocínio, eficiência computacional e aspectos arquiteturais; e o contexto (C) abrangeu pesquisas e aplicações de sistemas de recomendação. Segundo Kitchenham e Charters (2007), a definição das questões de pesquisa constitui uma etapa fundamental da revisão sistemática, pois orienta a investigação e determina as informações relevantes para a análise dos estudos.



Embora o planejamento original tenha priorizado os sistemas de recomendação, a síntese dos estudos selecionados foi posteriormente ampliada para examinar diferentes formas de integração arquitetural em outros domínios, distinguindo a combinação de mecanismos computacionais da integração entre agentes de paradigmas distintos. Também foi incorporada a questão complementar QC3.1, vinculada à QP3, para examinar especificamente os sistemas de recomendação. Essa ampliação constitui um desdobramento analítico posterior à seleção dos estudos e não implicou a realização de novas buscas. As questões de pesquisa e seus respectivos objetivos analíticos são apresentados na Tabela 1.

O procedimento de extração de dados teve como objetivo reunir as informações necessárias para responder às questões de pesquisa (KITCHENHAM; CHARTERS, 2007). A

análise considerou o problema investigado, as características das abordagens apresentadas, os mecanismos de integração, os domínios de aplicação, as evidências de avaliação e as limitações relatadas pelos estudos. Na síntese ampliada, buscou-se distinguir arquiteturas que integram agentes de paradigmas distintos daquelas que apenas combinam LLMs com outros mecanismos computacionais, sem presumir equivalência entre essas configurações.

A diretriz PRISMA foi utilizada como referência para organizar e relatar o processo de seleção dos estudos, contribuindo para a transparência das etapas de identificação, triagem e inclusão das publicações analisadas (PAGE et al., 2021). A busca primária foi conduzida nas plataformas IEEE Xplore e ScienceDirect (Elsevier), em 25 de junho de 2026. Para fins de reprodutibilidade, as consultas foram repetidas em 29 de setembro de 2026, com filtro de publicação entre 2021 e 2026, sendo as contagens desta segunda consulta apresentadas na Tabela 2. O processo de busca e seleção compreendeu as seguintes etapas:

Busca e recuperação: Foram utilizadas três strings de busca, uma para cada eixo temático, inseridas exatamente como segue, sem operadores booleanos: "LLM Recommender", "Classical Intelligent Agents" e "AI Hybrid Architecture". As mesmas strings foram aplicadas nas duas bases. Como não foram utilizados operadores, as bases trataram os termos de forma independente, o que resultou em um volume amplo de registros (Tabela 2) e tornou indispensável a triagem manual descrita seguidamente.



Triagem e processamento linguístico: Foram analisados os títulos e, em seguida, os resumos dos registros recuperados. Os resumos em inglês foram traduzidos para o português com o apoio de LLM (Gemini 3.6 Flash), exclusivamente para auxiliar na compreensão do conteúdo. A decisão de inclusão ou exclusão foi tomada manualmente pelos autores, com base nos critérios definidos abaixo. Nenhum modelo de LLM foi utilizado para decidir sobre a inclusão ou exclusão dos estudos. Ao final da leitura de títulos, 28 registros foram retidos e, após a leitura dos resumos, 20 estudos selecionados.

Aplicação dos critérios de inclusão e exclusão: A seleção dos estudos foi realizada mediante a aplicação sequencial dos critérios definidos para a revisão. Inicialmente, foi considerado o período de publicação, tendo como recorte temporal prioritário os trabalhos publicados entre 2021 e 2026. Em seguida, foi realizada a análise dos títulos, verificando sua aderência aos eixos temáticos da pesquisa e ao escopo delimitado pelo framework PICOC. Os trabalhos que apresentaram aderência nessa etapa foram submetidos à leitura dos respectivos resumos, com o objetivo de verificar sua pertinência em relação às questões de pesquisa e ao

escopo da revisão. Posteriormente, os estudos retidos foram submetidos a uma leitura exploratória, contemplando principalmente resumo, introdução e conclusão. Ao final desse processo, foram selecionados os estudos considerados aderentes ao contexto da pesquisa e aos critérios estabelecidos para a revisão sistemática. Os critérios e suas justificativas são apresentados a seguir.

Tabela 1 – Questões de pesquisa e respectivos objetivos analíticos

| ID | Questão de pesquisa | Objetivo analítico |
|---|---|---|
| QP1 | Como as arquiteturas híbridas integram os paradigmas de agentes clássicos e baseados em LLM? | Mapear os padrões de design arquitetural e fundamentar a taxonomia. |
| QP2 | Quais são as principais vantagens e limitações dessas abordagens híbridas frente aos modelos puros? | Analisar as vantagens, limitações e compromissos relatados nos estudos, considerando as evidências disponíveis. |
| QP3 | Em quais domínios de aplicação e cenários de dados essas arquiteturas estão sendo validadas? | Identificar o nível de maturidade e os principais contextos de uso prático na computação. |
| QC3.1 | Como os paradigmas de agentes investigados são empregados em sistemas de recomendação e quais limitações de integração são identificadas neste domínio? | Caracterizar as abordagens de recomendação presentes na amostra e examinar suas limitações arquiteturais. |
| QP4 | Quais são os principais desafios em aberto e lacunas de pesquisa nessa hibridização? | Identificar os principais desafios, limitações e oportunidades para pesquisas futuras sobre arquiteturas híbridas. |

Foram utilizados exatamente cinco critérios de avaliação para a inclusão dos artigos, a seguir.

CI 1. Escala Temporal: Publicação entre 2021 e 2026. Os LLMs generativos e os agentes construídos sobre eles se consolidaram a partir de 2022. O período delimitado cobre a evolução recente dos LLMs e das arquiteturas investigadas, evitando comparações entre abordagens de gerações tecnológicas muito distintas.



CI 2. Eixos Temáticos: Aderência a pelo menos um dos três eixos temáticos, ou seja, LLM aplicado à recomendação, agentes inteligentes anteriores aos LLMs (sistemas multiagentes, aprendizagem por reforço, controle) ou arquiteturas que combinam os dois paradigmas. Esse critério deriva diretamente da população, da intervenção e da comparação do PICOC.

CI 3. Descrição da arquitetura, do framework ou do método com detalhe suficiente para extrair problema, abordagem, métricas de desempenho e limitações, dimensões necessárias para responder às QP1, QP2 e QP4.

CI 4. Validação ou aplicação em um domínio identificável, condição para responder à QP3, que trata dos domínios e do nível de maturidade.

CI 5. Transferência arquitetural: Foram admitidos estudos de outros domínios (por exemplo, manufatura inteligente) quando apresentam padrões de integração entre agentes

clássicos e LLMs aplicáveis a sistemas de recomendação. A escassez de trabalhos que integram explicitamente os dois paradigmas em recomendação justifica esse alargamento, e a relevância desses estudos é avaliada pelo padrão arquitetural que documentam, e não pelo domínio de aplicação.

Foram utilizados cinco critérios de exclusão a seguir.

CE 1. Publicação fora do período de 2021 a 2026, pela justificativa do CI1.

CE 2. Ausência de conteúdo capaz de ser relevante no que se refere a responder às Questões de Pesquisa.

CE 3. Ausência de relação com agentes ou LLMs, como estudos de recomendação puramente baseados em filtragem colaborativa ou em modelos que não usam nenhum dos dois paradigmas.

CE 4. Uso de LLM ou de agente apenas como ferramenta auxiliar, sem descrição de sua integração na arquitetura, situação em que não é possível extrair o padrão de integração.

CE 5. Duplicidade, incluindo versões preliminares de um mesmo trabalho.

Os critérios apresentados correspondem ao protocolo originalmente empregado na seleção dos estudos e foram preservados para assegurar a rastreabilidade metodológica. A ampliação posterior da análise não modificou os critérios de seleção nem implicou a realização de novas buscas. Dessa forma, as conclusões sobre a integração entre agentes clássicos e agentes baseados em LLM em diferentes domínios devem ser interpretadas considerando a abrangência da amostra originalmente selecionada, sem pressupor cobertura exaustiva de todas as arquiteturas híbridas existentes na literatura.



Tabela 2 – Contagem de registros recuperados e estudos incluídos na revisão

| Strings | ScienceDirect (Elsevier) | IEEE Xplore | Total Bruto Encontrado | Incluídos na Planilha Final |
|---|---|---|---|---|
| 1. LLM Recommender | 10.383 | 563 | 10.946 | 9 |
| 2. Classical Intelligent Agents | 20.772 | 340 | 21.112 | 2 |
| 3. AI Hybrid Architecture | 79.132 | 5.692 | 84.824 | 9 |
| TOTAL | 110.287 | 6.595 | 116.882 | 20 |

Os quantitativos apresentados na Tabela 2 correspondem às ocorrências brutas relatadas para as consultas realizadas nas bases selecionadas, não representando necessariamente publicações únicas. Após a triagem por títulos e resumos, foram selecionados 20 estudos para a análise. Como não estão disponíveis registros individualizados das etapas intermediárias de seleção e de duplicação, não foi possível reconstruir integralmente o fluxo de identificação e

exclusão das publicações. Essa limitação deve ser considerada na interpretação da rastreabilidade e da reprodutibilidade do processo de revisão.

## 3 RESULTADOS

A análise dos 20 estudos selecionados foi realizada com o auxílio de traduções produzidas por modelo de linguagem (Gemini 3.6 Flash), acompanhadas de leitura parcial dos textos originais. Para cada estudo, foram sintetizadas informações relativas ao problema investigado, à abordagem proposta, às métricas ou evidências de avaliação e às limitações relatadas. Essas informações fundamentaram a caracterização dos trabalhos e a análise de suas contribuições para as questões de pesquisa.

Considerando a ampliação analítica descrita na metodologia, os estudos foram examinados quanto às diferentes formas de utilização e integração entre agentes inteligentes clássicos, agentes baseados em LLM e componentes computacionais associados. A análise buscou distinguir arquiteturas que integram agentes de paradigmas distintos daquelas que incorporam mecanismos clássicos e LLM no interior de agentes ou combinam modelos e ferramentas sem constituir agentes independentes. Para facilitar a identificação, os 20 estudos receberam códigos de A1 a A20, associados aos respectivos autores na Tabela 3 e utilizados na Tabela 4 e ao longo da análise dos resultados. A Tabela 3 apresenta a caracterização geral dos estudos e sua relação com as questões de pesquisa, enquanto a Tabela 4 sintetiza os aspectos arquiteturais, os domínios de aplicação e as formas de validação identificadas.



Dos 20 estudos analisados, seis são revisões de literatura (surveys), enquanto os demais apresentam cinco frameworks, três sistemas multi-agentes, quatro métodos e dois modelos. Essa distribuição reúne contribuições teóricas, arquiteturais e experimentais, permitindo examinar diferentes perspectivas sobre agentes inteligentes clássicos, agentes baseados em LLM e suas possibilidades de integração. Entretanto, os estudos apresentam níveis distintos de detalhamento e validação, exigindo que suas evidências sejam interpretadas conforme a natureza de cada contribuição.

Para caracterizar as estratégias arquiteturais identificadas, os estudos foram organizados em cinco categorias analíticas: integração interagentes (H1), hibridização intragente (H2), combinação parcial ou de componentes (H3), ausência de integração entre os paradigmas investigados (NH) e revisões secundárias (RS). A categoria H1 corresponde a arquiteturas nas quais agentes clássicos e agentes baseados em LLM desempenham funções próprias e interagem

no mesmo sistema. A categoria H2 caracteriza agentes que incorporam mecanismos clássicos e componentes baseados em LLM à sua estrutura funcional. A categoria H3 reúne combinações entre LLMs, agentes ou mecanismos computacionais que não demonstram integração efetiva entre agentes independentes dos dois paradigmas. A categoria NH compreende abordagens sem essa integração identificada, enquanto RS agrupa as revisões secundárias. Essas categorias foram elaboradas na etapa de síntese dos estudos selecionados e não constituíram critérios adicionais de seleção bibliográfica.

Tabela 3 – Caracterização dos 20 estudos analisados

| ID | Estudo | Contribuição | QPs respondidas |
|---|---|---|---|
| A1 | Nawara e Kashef (2025) | Survey | QP1(p), QP2(p), QP3, QP4 |
| A2 | Yang et al. (2024) | Framework | QP1(p), QP2, QP3, QP4 |
| A3 | Wesner et al. (2026) | MAS | QP1(p), QP2, QP3, QP4 |
| A4 | Shi et al. (2025) | Modelo | QP1(p), QP2, QP3, QP4 |
| A5 | Raza et al. (2026a) | Survey | QP2(p), QP3, QP4 |
| A6 | Zhao et al. (2026a) | MAS | QP1, QP2, QP3, QP4 |
| A7 | Nisa et al. (2026) | Survey | QP1(p), QP2(p), QP3(p), QP4 |
| A8 | Liu et al. (2026) | Método | QP1(p), QP2, QP3, QP4 |
| A9 | Raza et al. (2026b) | Survey | QP1(p), QP2(p), QP3(p), QP4 |
| A10 | Yang et al. (2026) | Método | QP1(p), QP2, QP3, QP4 |
| A11 | Gu et al. (2026) | MAS | QP1(p), QP2, QP3, QP4 |
| A12 | Pan et al. (2026) | Survey | QP2(p), QP4 |
| A13 | Xie et al. (2026) | Método | QP2, QP3, QP4 |
| A14 | Zhao et al. (2026b) | Survey | QP2(p), QP3, QP4 |
| A15 | Liu et al. (2025) | Framework | QP3, QP4 |
| A16 | Farahani, Khan e Wuest (2026) | Framework | QP1, QP2, QP3, QP4 |
| A17 | Ben Hassouna, Chaari e Belhaj (2026) | Framework | QP1(p), QP4 |
| A18 | Costa, Skogestad e Nogueira (2026) | Framework | QP3, QP4 |
| A19 | Hu et al. (2026) | Método | QP1(p), QP2, QP3, QP4 |
| A20 | Yue, Tong e Yu (2026) | Modelo | QP2, QP3, QP4 |

**Legenda:** MAS = sistema multiagente; QP = questão de pesquisa; (p) = contribuição parcial ou indireta.

Os seis surveys selecionados (A1, A5, A7, A9, A12 e A14) contribuem para contextualizar a evolução dos sistemas de recomendação baseados em LLM e da IA agêntica, além de identificar desafios recorrentes. A1 (Nawara e Kashef, 2025) discute abordagens de recomendação baseadas em LLM e limitações relacionadas à escalabilidade, à privacidade e às alucinações. A7 (Nisa et al., 2026) aborda desafios de avaliação, alinhamento e integração em sistemas multiagentes, enquanto A9 (Raza et al., 2026b) examina questões de confiança, risco, segurança e governança. Os trabalhos A5 (Raza et al., 2026a), A12 (Pan et al., 2026) e A14 (Zhao,

Y. et al., 2026) sintetizam avanços e limitações dos sistemas de recomendação, incluindo complexidade computacional, tratamento de informações heterogêneas e avaliação das abordagens.

Essas revisões contribuem principalmente para a identificação de limitações e desafios relacionados à QP4, oferecendo também subsídios conceituais para as demais questões. Entretanto, como estudos secundários, não constituem evidência direta de implementação ou validação das arquiteturas investigadas. Sua inclusão na síntese complementa a análise dos trabalhos primários, sem que seus resultados sejam interpretados como novas evidências experimentais.

Entre os cinco frameworks analisados (A2, A15, A16, A17 e A18), identificam-se diferentes formas de organização de agentes, modelos de linguagem e mecanismos computacionais especializados. Essas contribuições apresentam níveis distintos de integração arquitetural e abrangem aplicações em nutrição, manufatura, mobilidade, engenharia de software e controle de processos.

O estudo A2 (Yang et al., 2024) propõe o ChatDiet, uma arquitetura voltada à recomendação nutricional que combina LLMs com modelos causais e populacionais. Embora apresente integração entre componentes de diferentes abordagens, esses modelos especializados não constituem necessariamente agentes inteligentes clássicos autônomos, caracterizando uma combinação parcial de mecanismos (H3).



O estudo A16 (Farahani, Khan e Wuest, 2026) apresenta uma arquitetura para manutenção prescritiva em manufatura, combinando um agente planejador baseado em LLM com agentes especializados e mecanismos baseados em regras. Essa organização constitui o exemplo mais explícito de integração interagentes (H1) identificado na amostra. A arquitetura contempla funções de planejamento, percepção, pré-processamento, análise e otimização, com validação relatada por meio de prova de conceito utilizando dados industriais offline. Embora demonstre a viabilidade de execução coordenada das etapas, as evidências apresentadas não estabelecem superioridade geral em relação a outras configurações arquiteturais.

Os estudos A15 (Liu et al., 2025), A17 (Ben Hassouna, Chaari e Belhaj, 2026) e A18 (Costa, Skogestad e Nogueira, 2026) apresentam contribuições complementares. A15 investiga capacidades de agentes baseados em LLM em ambientes de mobilidade; A17 propõe um framework de modelagem de agentes LLM, incluindo diferentes organizações de componentes internos; e A18 apresenta uma arquitetura multiagente especializada em controle, sem

integração identificada com agentes baseados em LLM. Embora relevantes para compreender o projeto e a especialização funcional de agentes, esses estudos não demonstram a integração interagentes investigada na categoria H1.

Em relação às questões de pesquisa, A16 fornece a principal evidência arquitetural para a QP1 entre os frameworks, enquanto A2 e A17 contribuem parcialmente para a compreensão de diferentes formas de organização e combinação de componentes. A QP2 é contemplada pelas vantagens e limitações relatadas, incluindo aspectos de personalização, adaptabilidade, alucinação e latência, embora os estudos não forneçam uma comparação uniforme entre paradigmas. A QP3 evidencia aplicações em diferentes domínios e níveis de validação, enquanto a QP4 reúne desafios relacionados à confiabilidade, à coordenação, à governança e à avaliação das arquiteturas.

Os quatro estudos classificados como métodos (A8, A10, A13 e A19) apresentam abordagens voltadas a sistemas de recomendação, explorando diferentes formas de utilização de modelos de linguagem e mecanismos de aprendizagem, avaliação ou recuperação de informações.

Tabela 4 – Caracterização dos paradigmas e padrões arquiteturais identificados

| ID | Contribuição | Agente clássico | Agente LLM | Padrão | Rec. | Domínio | Validação |
|---|---|---|---|---|---|---|---|
| A1 | Survey | - | - | RS | Sim | Recomendação | Revisão |
| A2 | Framework | Não | Parcial | H3 | Sim | Nutrição | Caso único |
| A3 | MAS | Não identificado | Sim | H2 | Não | Manufatura | Sintética |
| A4 | Modelo | Não | Não | H3 | Sim | E-commerce | Offline + caso real |
| A5 | Survey | - | - | RS | Sim | Recomendação | Revisão |
| A6 | MAS | Sim* | Sim* | H2 | Não | Manufatura | Laboratório |
| A7 | Survey | - | - | RS | Não | IA agêntica | Revisão |
| A8 | Método | Sim | Não | H3 | Sim | Vídeos curtos | Offline |
| A9 | Survey | - | - | RS | Não | IA agêntica | Revisão |
| A10 | Método | Não | Não | H3 | Sim | Filmes/livros | Offline |
| A11 | MAS | Sim | Não | H3 | Não | Manufatura | Protótipo |
| A12 | Survey | - | - | RS | Sim | Rec. multimodal | Revisão |
| A13 | Método | Não | Não | NH | Sim | Mídia | Offline |
| A14 | Survey | - | - | RS | Sim | Rec. generativa | Revisão |
| A15 | Framework | Não | Sim | NH | Não | Mobilidade | Estudo de caso |
| A16 | Framework | Sim | Sim | H1 | Não | Manufatura | Prova de conceito |
| A17 | Framework | Não | Sim | NH | Não | Eng. de software | Qualitativa |
| A18 | Framework | Parcial | Não | NH | Não | Controle de processos | Estudo de caso |
| A19 | Método | Não | Não | H3 | Sim | Filmes/livros | Offline |

| ID | Contribuição | Agente clássico | Agente LLM | Padrão | Rec. | Domínio | Validação |
|---|---|---|---|---|---|---|---|
| A20 | Modelo | Sim | Não | NH | Sim | Vídeos de dança | Exp. controlado |

**Legenda:** H1 = integração entre agentes clássicos e LLM; H2 = hibridização de mecanismos no mesmo agente; H3 = combinação parcial de componentes; NH = sem integração; RS = revisão secundária; Rec. = sistemas de recomendação; - = não aplicável. Em A3 e A6, H2 indica mecanismos clássicos e LLM combinados nos agentes, sem comprovação de agentes independentes dos dois paradigmas.

O estudo A8 (Liu et al., 2026) apresenta o DRUM, um método de aprendizagem por reforço offline para recomendação interativa, no qual um LLM ajustado com LoRA atua como modelo de usuário e fornece informações utilizadas na avaliação das ações de um agente de aprendizagem por reforço. Embora a abordagem combine um agente clássico com um LLM, este desempenha a função de simulador, não sendo caracterizado como agente independente. Assim, o estudo constitui um exemplo de combinação parcial (H3).

O estudo A10 (Yang et al., 2026) propõe o Critic-LLM-RS, que associa um LLM a um modelo crítico independente de aprendizagem de máquina, utilizando feedback para refinar recomendações. Essa combinação caracteriza a integração funcional entre modelos, mas não demonstra a presença de um agente clássico. De maneira semelhante, A19 (Hu et al., 2026) apresenta o ReACT, método baseado na recuperação de informações para reduzir ruídos do histórico de interações antes do processamento pelo LLM. Ambos foram classificados como combinações de componentes (H3).



Por sua vez, A13 (Xie et al., 2026) investiga uma abordagem de recomendação apoiada em LLM e seleção contextual de exemplos por aprendizagem contrastiva. Embora incorpore mecanismos especializados de processamento, o estudo não apresenta integração entre agentes clássicos e agentes baseados em LLM, sendo classificado como NH para o objeto específico desta revisão.

Em relação às questões de pesquisa, A8, A10 e A19 contribuem parcialmente para a QP1, ao demonstrar combinações entre agentes, modelos de linguagem e mecanismos auxiliares. A QP2 é contemplada pelos resultados e limitações relatados nos quatro estudos, enquanto a QP3 compreende aplicações em recomendação de vídeos, filmes, livros e outros conteúdos. Quanto à QP4, destacam-se limitações relacionadas à avaliação offline, ao custo computacional, ao tratamento do histórico de interações, à personalização e à ausência de avaliação em condições de operação contínua.

Esses trabalhos também oferecem evidências para a análise complementar dos sistemas de recomendação, demonstrando diferentes formas de incorporação de LLMs e mecanismos

clássicos. Entretanto, as combinações identificadas não devem ser interpretadas automaticamente como arquiteturas compostas por agentes independentes dos dois paradigmas.

Os três estudos classificados como sistemas multiagentes (A3, A6 e A11) apresentam aplicações em manufatura e diferentes formas de combinação entre mecanismos clássicos de processamento, aprendizagem e componentes baseados em LLM.

O estudo A3 (Wesner et al., 2026) propõe um sistema multiagente para detecção automatizada de falhas em processos de inspeção na remanufatura. A arquitetura utiliza agentes especializados que executam procedimentos clássicos de extração de características de sinais e empregam um LLM para estimar probabilidades de falhas e produzir explicações diagnósticas. Essa organização caracteriza uma forma de hibridização funcional intragente (H2), pois combina mecanismos clássicos e capacidades baseadas em LLM na estrutura dos agentes, sem demonstrar a cooperação entre agentes clássicos e agentes LLM independentes. A avaliação foi realizada com dados sintéticos, apresentando limitações especialmente em situações com múltiplas falhas.

O estudo A6 (Zhao et al. 2026a), apresenta um sistema multiagente para manufatura inteligente no qual os agentes incorporam mecanismos de negociação por licitação, regras heurísticas e módulos de decisão baseados em LLM. Essa organização também foi caracterizada como hibridização intragente (H2), embora apresente mecanismos de negociação e coordenação próprios do sistema multiagente. Os autores relatam validação em ambiente físico laboratorial, com resultados relacionados ao desempenho e à estabilidade do escalonamento. Entretanto, a presença de mecanismos clássicos e LLM no interior dos agentes não demonstra, isoladamente, a integração entre dois agentes autônomos de paradigmas distintos.



Por sua vez, A11 (Gu et al., 2026) apresenta uma abordagem de escalonamento industrial que combina aprendizagem por reforço multiagente com apoio de LLM. Nesse caso, os agentes de aprendizagem por reforço participam do processo decisório, enquanto o LLM atua como componente auxiliar. Como não foi demonstrada autonomia agêntica independente do componente LLM, o estudo foi classificado como combinação parcial (H3).

Esses três trabalhos contribuem para a QP1 ao apresentar diferentes formas de organização funcional de mecanismos clássicos e baseados em LLM. A QP2 é contemplada pelos resultados relacionados ao diagnóstico, ao escalonamento e às limitações computacionais relatadas. A QP3 evidencia aplicações industriais com avaliações conduzidas em dados sintéticos, ambiente laboratorial e protótipo, respectivamente. Quanto à QP4, destacam-se

desafios relacionados à robustez, à latência, à dependência de mecanismos de referência e à generalização dos resultados para ambientes operacionais distintos.

Em conjunto, os estudos demonstram que a utilização de sistemas multiagentes associados a LLMs não implica necessariamente a existência de uma arquitetura de integração interagentes. A diferenciação entre mecanismos incorporados aos agentes e agentes independentes constitui, portanto, um aspecto relevante para caracterizar as estratégias arquiteturais identificadas na revisão.

Os dois estudos classificados como modelos (A4 e A20) apresentam abordagens de recomendação que combinam diferentes técnicas de Inteligência Artificial, mas não demonstram integração entre agentes clássicos e agentes baseados em LLM independentes.

O estudo A4 (Shi et al., 2025) propõe o LLM-SRR, combinando um modelo de linguagem com mecanismos de raciocínio baseados em grafos de conhecimento. O LLM é utilizado para extrair informações e produzir explicações, enquanto o componente de raciocínio por subgrafos participa da classificação e ordenação dos itens recomendados. A avaliação relatada contempla conjuntos de dados da Amazon e um caso de venda cruzada. Embora combine mecanismos semânticos e estruturados, a arquitetura não caracteriza necessariamente a integração entre agentes de paradigmas distintos, sendo classificada como H3.



O estudo A20 (Yue, Tong e Yu, 2026) apresenta uma abordagem para recomendação de vídeos de dança, combinando aprendizagem profunda e aprendizagem por reforço. A arquitetura não utiliza LLM e, portanto, não representa uma integração entre os paradigmas investigados. Entretanto, constitui uma referência complementar para compreender o emprego de mecanismos clássicos de decisão em aplicações de recomendação, tendo sido classificada como NH.

Ambos contribuem para a QP2 e a QP3 ao apresentar resultados de avaliação e contextos de aplicação, além de fornecer informações sobre limitações relevantes para a QP4. A contribuição de A4 para a QP1 é parcial e está relacionada à combinação funcional de mecanismos de processamento semântico e raciocínio estruturado, enquanto A20 oferece um contraponto baseado em técnicas que não envolvem LLMs.

Considerando a amostra analisada, os resultados evidenciam diferentes níveis de integração entre mecanismos clássicos e componentes baseados em LLM. A16 constitui o exemplo mais explícito de integração entre agentes de paradigmas distintos (H1), enquanto A3 e A6 apresentam hibridização funcional intragente (H2). Outros estudos combinam agentes,

modelos de linguagem e mecanismos especializados sem demonstrar a integração entre agentes independentes dos dois paradigmas (H3). Essa heterogeneidade evidencia que a caracterização das arquiteturas depende não apenas das tecnologias utilizadas, mas também dos papéis funcionais, dos mecanismos de decisão e das formas de interação entre os componentes. As evidências de avaliação também apresentam diferenças quanto aos domínios, às métricas e às condições experimentais, limitando comparações diretas entre as abordagens.

De maneira complementar, a análise dos sistemas de recomendação identificou abordagens baseadas em agentes clássicos, LLMs e combinações de componentes, incluindo mecanismos de aprendizagem por reforço, raciocínio estruturado e recuperação de informações. Entretanto, entre os estudos selecionados, não foi identificada uma arquitetura que demonstrasse simultaneamente a presença de agentes clássicos e agentes baseados em LLM independentes, integrados em um sistema de recomendação no sentido considerado nesta revisão. Essa observação caracteriza uma particularidade da amostra analisada e não constitui evidência suficiente para afirmar a inexistência dessa integração em toda a literatura científica.

## 4 DISCUSSÃO

Neste estudo, foram analisados 20 trabalhos científicos com o objetivo de caracterizar diferentes formas de integração entre agentes inteligentes clássicos e agentes baseados em grandes modelos de linguagem (LLMs), considerando seus padrões arquiteturais, domínios de aplicação, vantagens, limitações e desafios. A classificação apresentada nas Tabelas 3 e 4 permitiu organizar as contribuições selecionadas e estabelecer distinções entre arquiteturas que integram agentes de paradigmas distintos, abordagens que combinam mecanismos clássicos e componentes baseados em LLM no interior de agentes e sistemas que utilizam esses recursos de maneira complementar, sem caracterizar uma integração interagentes.



A análise evidenciou diferentes níveis de hibridização. O estudo A16 apresentou a configuração mais explícita de integração interagentes (H1), combinando um agente planejador baseado em LLM com agentes especializados e mecanismos estruturados de decisão no contexto da manutenção prescritiva. Por sua vez, A3 e A6 foram classificados como exemplos de hibridização intragente (H2), nos quais mecanismos clássicos de processamento, negociação ou decisão são combinados com capacidades baseadas em LLM na estrutura funcional dos agentes. Outros trabalhos, como A2, A8, A10 e A11, apresentaram combinações parciais entre agentes, modelos de linguagem e componentes especializados (H3), sem demonstrar necessariamente a

existência de agentes independentes dos dois paradigmas. Essas diferenças indicam que a presença simultânea de técnicas clássicas e LLMs não é suficiente, isoladamente, para caracterizar uma arquitetura de integração entre agentes inteligentes.

As quatro questões principais de pesquisa permitiram identificar características e limitações relevantes das abordagens analisadas. Em relação à QP1, foram observados diferentes padrões de organização e interação entre componentes, desde a hibridização funcional intragente até a coordenação de agentes especializados. Quanto à QP2, os estudos relatam benefícios associados ao desempenho, à personalização, à interpretação de informações e à adaptabilidade, embora essas evidências sejam heterogêneas e não permitam estabelecer superioridade geral das arquiteturas híbridas em relação às abordagens individuais. Na QP3, verificaram-se aplicações em diferentes domínios, incluindo manufatura, mobilidade, controle de processos e recomendação, com avaliações conduzidas em condições experimentais distintas, frequentemente restritas a dados offline, simulações, protótipos ou provas de conceito. Por fim, a QP4 evidenciou desafios relacionados à confiabilidade, às alucinações, à latência, ao custo computacional, à escalabilidade, à privacidade e à necessidade de avaliações comparativas e protocolos de integração mais consistentes.



De maneira complementar, a análise dos sistemas de recomendação (QC3.1) identificou abordagens que utilizam agentes clássicos, LLMs e diferentes combinações de componentes. Entretanto, entre os estudos selecionados, não foi identificada uma arquitetura que demonstrasse simultaneamente a presença de agentes clássicos e agentes baseados em LLM independentes, integrados no contexto da recomendação personalizada de itens. Essa observação sugere uma oportunidade de investigação nesse domínio específico, mas não permite concluir que tais arquiteturas estejam ausentes de toda a literatura científica.

A interpretação dos resultados também deve considerar as limitações metodológicas da revisão. A seleção original dos estudos priorizou sistemas de recomendação, embora tenha admitido trabalhos de outros domínios por sua relevância arquitetural. Portanto, a ampliação analítica realizada permite caracterizar diferentes formas de integração presentes na amostra, mas não assegura cobertura exaustiva das arquiteturas híbridas existentes em todos os contextos computacionais. Além disso, a diversidade dos tipos de contribuição, das métricas e dos procedimentos de avaliação limita comparações diretas entre os estudos. A indisponibilidade de registros individualizados da triagem também limita a reconstrução integral do processo de seleção dos estudos. Nesse cenário, os resultados indicam a importância de aprofundar a

investigação dos mecanismos de coordenação entre agentes, das responsabilidades atribuídas a cada paradigma e das condições em que sua integração pode oferecer benefícios verificáveis.

## 5 CONCLUSÃO

Este estudo apresentou uma revisão sistemática da literatura sobre a integração entre agentes inteligentes clássicos e agentes baseados em grandes modelos de linguagem (LLMs), a partir da análise de 20 estudos científicos. A principal contribuição consiste na caracterização de diferentes formas de hibridização, distinguindo arquiteturas que integram agentes de paradigmas distintos, abordagens que combinam mecanismos clássicos e componentes baseados em LLM no interior de agentes e sistemas que utilizam essas tecnologias de maneira complementar, sem necessariamente estabelecer uma integração entre agentes independentes.

Os resultados evidenciaram diferentes padrões arquiteturais, destacando-se a integração interagentes identificada em A16 e as formas de hibridização intragente observadas em A3 e A6. Embora os estudos relatem benefícios relacionados ao desempenho, à adaptabilidade e à utilização de capacidades complementares, as evidências disponíveis apresentam limitações quanto à abrangência das validações e à comparabilidade dos resultados. A análise complementar dos sistemas de recomendação também identificou oportunidades de investigação, uma vez que não foram encontradas, entre os estudos selecionados, arquiteturas que integrassem diretamente agentes clássicos e agentes baseados em LLM independentes nesse domínio.



Como perspectivas futuras, destacam-se a investigação de mecanismos de coordenação e comunicação entre agentes de diferentes paradigmas, o desenvolvimento de critérios mais precisos para caracterizar arquiteturas híbridas e a realização de avaliações experimentais que permitam examinar os benefícios e as limitações dessas integrações em diferentes contextos de aplicação. Considerando que a seleção original dos estudos priorizou sistemas de recomendação, pesquisas posteriores também poderão ampliar a cobertura bibliográfica para outros domínios, contribuindo para uma compreensão mais abrangente dos padrões arquiteturais e de sua aplicabilidade.

## REFERÊNCIAS

BEN HASSOUNA, A.; CHAARI, H.; BELHAJ, I. LLM-Agent-UMF: LLM-based agent unified modeling framework for seamless design of multi active/passive core-agent architectures. *Information Fusion*, v. 127, art. 103865, 2026. DOI: https://doi.org/10.1016/j.inffus.2025.103865.

COSTA, E. A.; SKOGESTAD, S.; NOGUEIRA, I. B. R. MAASC - Multi-agentic architecture specialised for control. *Journal of Process Control*, v. 162, art. 103702, 2026. DOI: https://doi.org/10.1016/j.jprocont.2026.103702.

FARAHANI, M. A.; KHAN, M. I.; WUEST, T. Hybrid agentic AI and multi-agent systems in smart manufacturing. *Journal of Manufacturing Systems*, v. 86, p. 612–623, 2026. DOI: https://doi.org/10.1016/j.jmsy.2026.04.002.

GU, W. et al. Large language model-empowered dynamic scheduling for intelligent hybrid flow shop using multi-agent deep reinforcement learning. *Advanced Engineering Informatics*, v. 71, art. 104294, 2026. DOI: https://doi.org/10.1016/j.aei.2025.104294.

HU, Z. et al. Retrieval-enhanced, adaptively collaborative, and temporal-aware user behavior comprehension for LLM-based sequential recommendation. *Information Processing & Management*, v. 63, n. 1, art. 104354, 2026. DOI: https://doi.org/10.1016/j.ipm.2025.104354.

KITCHENHAM, B.; CHARTERS, S. *Guidelines for performing systematic literature reviews in software engineering*. Version 2.3. EBSE Technical Report EBSE-2007-01, 2007.

LIU, E. et al. Human-AI agent interaction design towards agentic cockpits for intelligent vehicles: framework and a case study. *IFAC-PapersOnLine*, v. 59, n. 35, p. 573–578, 2025. DOI: https://doi.org/10.1016/j.ifacol.2025.12.539.

LIU, H. et al. DRUM: distributionally robust optimization for LLM-based user model in interactive recommendation. *Information Processing & Management*, v. 63, n. 6, art. 104746, 2026. DOI: https://doi.org/10.1016/j.ipm.2026.104746.



NAWARA, D.; KASHEF, R. A comprehensive survey on LLM-powered recommender systems: from discriminative, generative to multi-modal paradigms. *IEEE Access*, v. 13, p. 145772–145798, 2025. DOI: https://doi.org/10.1109/ACCESS.2025.3599832.

NISA, U.; SHIRAZI, M.; SAIP, M. A.; MOHD POZI, M. S. Agentic AI: the age of reasoning - a review. *Journal of Automation and Intelligence*, v. 5, n. 1, p. 69–89, 2026. DOI: https://doi.org/10.1016/j.jai.2025.08.003.

PAGE, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, v. 372, art. n71, 2021. DOI: https://doi.org/10.1136/bmj.n71.

PAN, L.; PAN, Z.; CAI, F.; CHEN, H. Multimodal recommender systems: a survey of representation, modeling, and optimization. *Information Fusion*, v. 128, art. 103991, 2026. DOI: https://doi.org/10.1016/j.inffus.2025.103991.

RAZA, S. et al. A comprehensive review of recommender systems: transitioning from theory to practice. *Computer Science Review*, v. 59, art. 100849, 2026a. DOI: https://doi.org/10.1016/j.cosrev.2025.100849.

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*, v. 7, p. 71–95, 2026b. DOI: https://doi.org/10.1016/j.aiopen.2026.02.006.

RUSSELL, S.; NORVIG, P. *Artificial intelligence: a modern approach*. 3. ed. Pearson, 2013.

SHI, G. et al. LLM-powered explanations: unraveling recommendations through subgraph reasoning. *Knowledge-Based Systems*, v. 329, art. 114307, 2025. DOI: https://doi.org/10.1016/j.knosys.2025.114307.

WANG, L. et al. A survey on large language model based autonomous agents. *Frontiers of Computer Science*, v. 18, art. 186345, 2024. DOI: https://doi.org/10.1007/s11704-024-40231-1.

WESNER, D.; KOCH, D.; MAS, V.; MATTHIESEN, S.; STAMER, F.; LANZA, G. LLM-powered multi-agent system for automated error detection in remanufacturing inspection. *Procedia CIRP*, v. 139, p. 174–178, 2026. DOI: https://doi.org/10.1016/j.procir.2025.09.022.

XI, Z. et al. The rise and potential of large language model based agents: a survey. *Science China Information Sciences*, v. 68, art. 121101, 2025. DOI: https://doi.org/10.1007/s11432-024-4222-0.

XIE, J. et al. Exploring prompt engineering with instructive shots to enhance LLMs as a recommender system. *Information Sciences*, v. 735, art. 123064, 2026. DOI: https://doi.org/10.1016/j.ins.2025.123064.

YANG, Z. et al. ChatDiet: empowering personalized nutrition-oriented food recommender chatbots through an LLM-augmented framework. *Smart Health*, v. 32, art. 100465, 2024. DOI: https://doi.org/10.1016/j.smhl.2024.100465.

YANG, Z.; XU, X.; DENG, K.; LI, L. Wise recommender: LLMs refined by iterative critics. *Information and Software Technology*, v. 192, art. 108021, 2026. DOI: https://doi.org/10.1016/j.infsof.2026.108021.



YUE, F.; TONG, J.; YU, Y. A case study on intelligent recognition and personalized recommendation of China classical dance style based on deep learning. *Journal of Cases on Information Technology*, v. 28, n. 1, p. 1–21, 2026. DOI: https://doi.org/10.4018/JCIT.411904.

ZHAO, Z. et al. A large language model-based multi-agent manufacturing system for intelligent shopfloors. *Advanced Engineering Informatics*, v. 69, art. 103888, 2026a. DOI: https://doi.org/10.1016/j.aei.2025.103888.

ZHAO, Y. et al. Generative recommender systems: a comprehensive survey on model, framework, and application. *Information Fusion*, v. 127, art. 103919, 2026b. DOI: https://doi.org/10.1016/j.inffus.2025.103919.