最新综述 | 大模型智能体时代的推荐系统

ML_RSer 机器学习与推荐算法 2025年01月22日 11:17 北京

TLDR: 本文全面概述了基于大模型的AI智能体（LLM Agent）与推荐系统互利共生的现状及未来发展方向，分析了LLM智能体模块（如用户画像、记忆、规划、行为模块和多智能体协作）对推荐系统的支持，以及推荐系统（如记忆、规划、工具和智能体推荐）如何优化智能体的运作。本文还深入探讨了二者在安全性、可解释性、公平性和隐私保护等方面的关键问题，并展望了当前的挑战和机遇，为未来研究提供了方向性指导。

Recommender Systems in the Era of Large Language Model Agents: A Survey

XI ZHU*, Rutgers University, United States
YU WANG*†, Netflix, United States
HANG GAO*, Rutgers University, United States
WUJIANG XU*, Rutgers University, United States
CHEN WANG, University of Illinois Chicago, United States
ZHIWEI LIU, Salesforce AI Research, United States
KUN WANG, Squirrel Ai Learning, United States
MINGYU JIN, Rutgers University, United States
LINSEY PANG, Salesforce, United States
QINGSONG WEN, Squirrel Ai Learning, United States
PHILIP S. YU, University of Illinois Chicago, United States
YONGFENG ZHANG, Rutgers University, United States

论文：https://bit.ly/3E0sp5M
仓库：https://github.com/agiresearch/AgentRecSys

摘要

近年来，大语言模型（LLMs）与推荐系统（RS）的融合开启了个性化和智能化用户体验的新时代。本综述全面概述了基于LLM的AI智能体（LLM Agent）与推荐系统协作的现状及未来发展方向。

我们首先分析了LLM智能体与推荐系统之间的互利共生关系。具体来说，我们从设计LLM智能体的各模块入手，分析了用户画像、记忆模块、规划模块、行为模块以及多智能体协作等关键组件在推荐系统中的应用。同时，我们探讨了推荐系统如何优化LLM智能体，重点关注记忆推荐、规划推荐、工具推荐和智能体推荐等任务，以及个性化LLM和LLM智能体的设计方法。

此外，我们深入探讨了可信智能体和推荐系统在安全性、可解释性、公平性和隐私保护等方面的关键问题。最后，综述展望了未来研究方向，提出了AI智能体与推荐系统交叉领域的新兴趋势和研究机遇，为这一快速发展的领域提供了深刻洞见与前瞻性指导。

研究动机

在LLM时代，推荐系统和智能体迎来了革命性的发展。一方面，由LLM支撑的智能体具备理解、规划、推理、解释和执行等能力，逐步演化为更强大、更通用的问题解决者。另一方面，LLM支持推荐系统的全流程，通过理解、生成、推理和解释等方式，更高效地增强用户建模与关键信息过滤。这表明，推荐系统与智能体在LLM时代展现了诸多共通点。

在LLM的支持下，智能体和推荐系统能够深度整合思想、原理与技术。具体而言，LLM智能体通过用户画像、记忆、规划和行动等模块，有效提升推荐系统的性能。而推荐系统则根据不同应用场景衍生出记忆推荐、规划推荐、工具推荐以及智能体推荐等任务，进一步优化智能体的决策过程。

本综述深入分析了智能体与推荐系统之间的协同关系，探讨了LLM智能体的各组成模块如何支持推荐系统的全流程，以及推荐系统如何反哺智能体的决策优化。我们全面展示了二者在LLM驱动下的特点和优势，并阐明了它们如何互利共生增强整体能力。

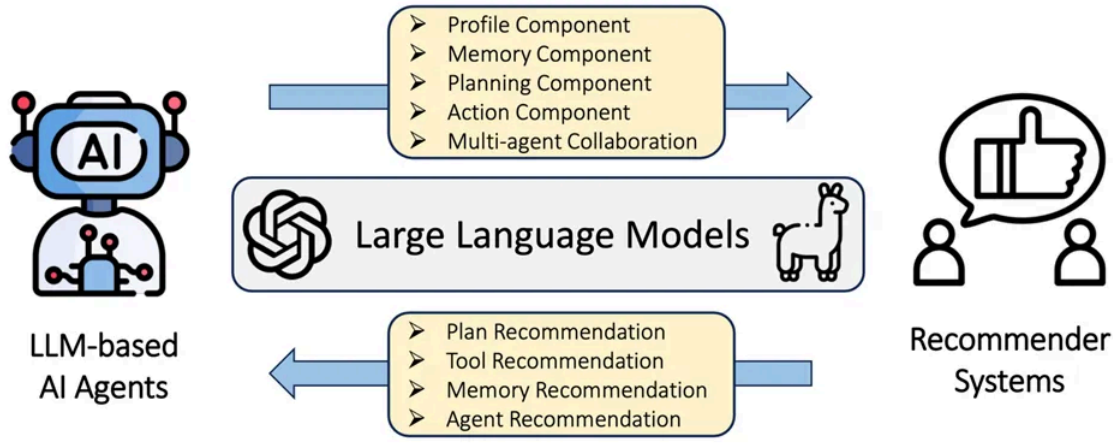

Fig. 2. The bidirectional relationship between AI agents and recommender systems in the era of LLMs.

面向推荐系统的大模型智能体

传统推荐系统在适应用户需求变化、处理复杂交互以及提升泛化能力方面存在不足。LLM智能体能够通过优化用户画像建模、积累先验知识、增强决策能力和提高任务执行效率，有效提升推荐系统的智能性与灵活性。通过精心设计智能体的画像、记忆、规划和行动模块，实现了个性化推荐与高效任务处理。

- **画像模块 (Profile Component)。** 用户画像对于确保推荐结果与用户偏好保持一致至关重要。因此，画像模块定义并封装了用户和物品的关键特征（traits），支持智能体模拟用户-物品交互行为，从而增强推荐的个性化和相关性。此外，还可以针对具体任务，通过智能体角色指令（agent role instructions）定义更多专门化角色，让多智能体之间交互与合作完成更复杂的任务。

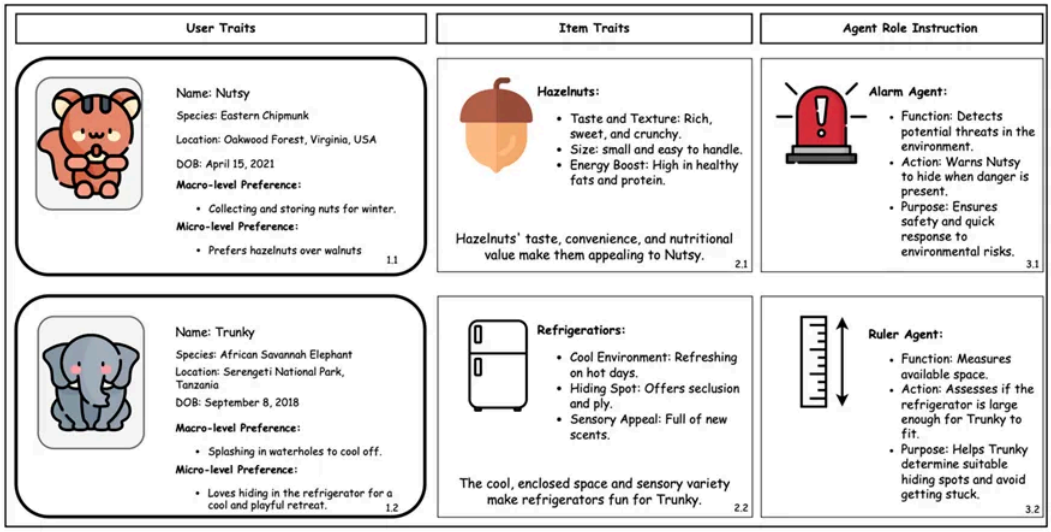

Fig. 4. Illustration of the profile component in LLM agents using the example of a squirrel and an elephant. The figure highlights how user traits, item traits, and agent role instructions function within the profile component. For user traits, the squirrel (Nutsy) demonstrates macro-level traits such as collecting and storing nuts and micro-level preferences like favoring hazelnuts over walnuts. The elephant (Trunky) displays macro-level behaviors such as socializing and cooling off, with micro-level preferences like hiding in a refrigerator. The item traits are represented through adaptive engagement properties that adjust to user needs. Agent role instructions are illustrated with the "alarm agent" for Nutsy, which detects threats and signals her to hide, and the "ruler agent" for Trunky, which measures whether a refrigerator is large enough for him to fit.

- **记忆模块 (Memory Component)。** 记忆模块使智能体能够记录先前的交互、用户偏好和环境信息，为上下文感知推荐和长期推荐奠定基础。从结构上，长短期记忆构建了分层记忆体系；从功能上，个性化记忆（personalized memory）与持久记忆（persistent memory）相结合，平衡了推荐中对个性化与通用知识的需求。同时，通过实时记忆（real-time memory）与反思记忆（reflective memory），智能体能够动态调整策略以适应环境变化。在多智能体系统中，协同记忆（collaborative memory）促进了智能体之间的信息共享与协作学习。综上，记忆模块是推荐系统与LLM智能体协同工作的核心组成部分。

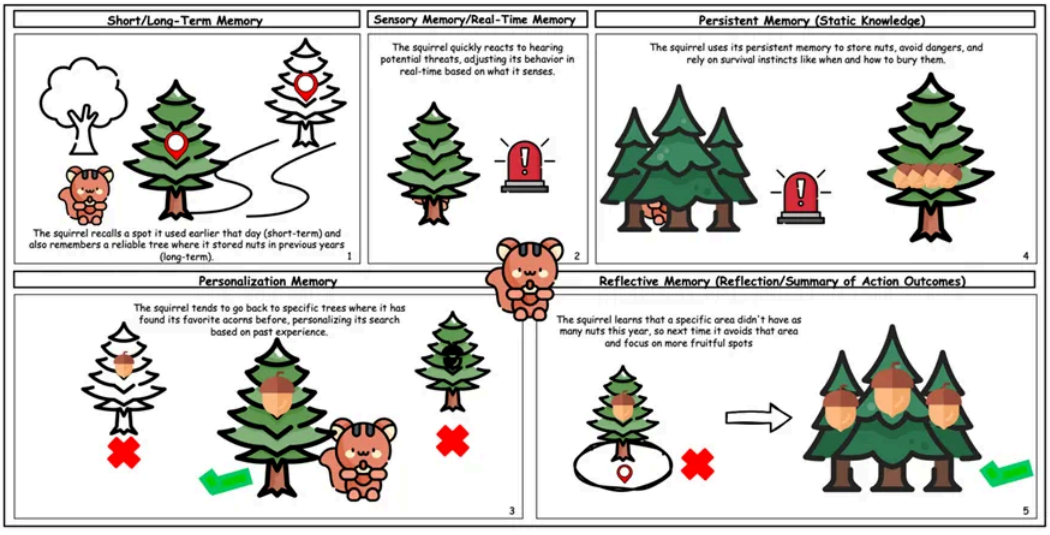

Fig. 5. Illustration of different memory types for LLM agents using the *"squirrel storing nuts"* example. This figure illustrates the different memory types utilized by the squirrel: short-term memory recalls recent hiding spots, while long-term memory retains knowledge of reliable locations used in previous years. Sensory memory processes immediate inputs, such as detecting nearby dangers, while personalization memory guides preferences for specific nuts or trees. Persistent memory stores static knowledge, including when and where to bury nuts and avoid threats. Finally, reflective memory enables the squirrel to adapt its foraging strategy based on past outcomes, enhancing its ability to make more informed decisions over time.

- **规划模块 (Planning Component)。** 规划模块将复杂的大任务分解为可管理的小步骤，生成并优化规划决策，从而支持LLM智能体高效地实现目标。以“大象放入冰箱”的示例为例，规划策略可

根据应用需求划分为静态规划（static planning）、响应式规划（reactive planning）、前瞻式规划（preactive planning）和反思式规划（reflective planning）等类型。规划模块是构建复杂高效推荐系统的重要模块与关键支撑。

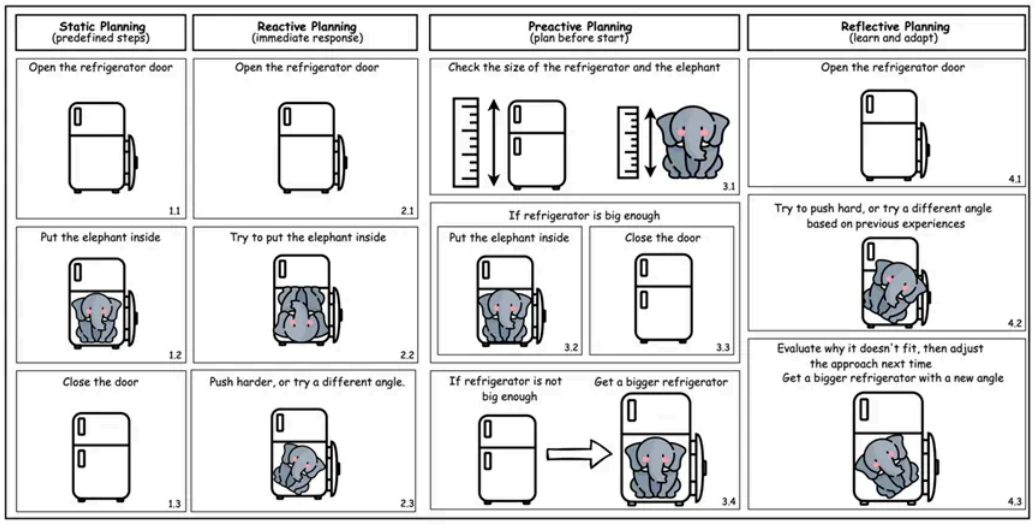

Fig. 6. Illustration of different planning strategies for LLM agents using the *"putting an elephant into a refrigerator"* example. The figure shows how static planning follows a fixed path, reactive planning responds to immediate stimuli, preactive planning anticipates possible obstacles, and reflective planning adapts over time based on past experience.

- **行为模块 (Action Component)。** 行动模块是智能体独有的核心组件，使其区别于一般的基于LLM的推荐系统。由规划模块的决策触发，行动模块支持智能体与环境、记忆模块及外部工具交互，并将处理结果反馈给智能体。在推荐系统中，行动根据功能划分为用户模拟行动（user simulation actions）、记忆相关行动（memory actions）和工具执行行动（tool execution actions）三大类别。

- **多智能体协作 (Multi-agent Collaboration)。** 多智能体协作通过智能体间的协同与任务分工，有效提升系统的综合能力。协作形式可分为同功能智能体的协作（如 RecAgent 和 Agent4Rec）以及不同功能智能体的分工协作（如 AgentCF 和 MACRec）。多智能体协作模式支持动态交互与知识共享，使系统能够应对更复杂的用户需求和推荐场景。

服务于大模型智能体的推荐系统

当前的 LLM 智能体在记忆保留、任务分解、工具选择和领域适应性等方面存在一定局限。引入推荐系统为解决这些问题提供了有效路径，通过优化记忆、规划、工具以及智能体自身，不仅提升了推荐的个性化，还支持动态选择最适合的模型或资源，从而增强系统的可扩展性和灵活性。推荐系统与 LLM 智能体的深度融合为智能体的未来发展和实际应用开辟了广阔前景。

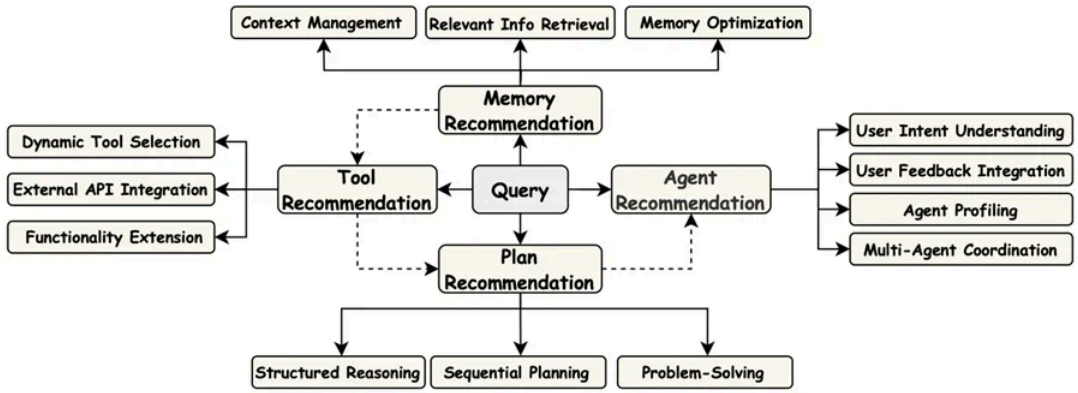

Fig. 8. An overview of how recommender systems enhance LLM agents. Memory, tool, plan, and agent recommendations can be viewed as a progressive framework that addresses problems from the simplest to increasingly complex levels.

- **记忆推荐 (Memory Recommendation)。** 记忆推荐通过动态选择和检索与当前任务相关的记忆，扩展 LLM 智能体的上下文能力，优化任务响应和决策效率。关键技术包括高效检索机制、神经网络架构优化以及结构化记忆的引入。

- **规划推荐 (Planning Recommendation)。** 规划推荐通过为 LLM 智能体提供结构化步骤和战略提示，缓解多步骤推理和复杂问题解决的局限性，提升任务管理能力、逻辑流畅性和一致性。关键方法包括内部推理增强（如 CoT 提示、自一致性方法）和外部计算与交互推理（如外部记忆存储、PAL 方法、ToT 提示）。

- **工具推荐 (Tool Recommendation)。** 工具推荐通过动态选择和使用专门的工具或 API，帮助 LLM 智能体在无法独立完成任务时借助外部资源实现目标。这一方法扩展了 LLM 智能体的功能，使其能够执行超越语言理解与生成范围的任务，例如客户服务、商业分析和决策支持。主要技术包括提示策略、结构化策略和检索策略。

- **智能体推荐 (Agent Recommendation)。** 从具体模块到 LLM 智能体本身，智能体推荐通过分析用户查询、理解用户意图并匹配智能体功能，提升用户体验和工作效率，确保智能体在特定领域（如代码开发、医疗诊断、法律分析和客户支持等）得到最佳应用。该领域仍处于初步发展阶段，主要包含意图理解和自适应推荐两种方法。

- **个性化大模型和智能体 (Personalized LLM and LLM Agents)。** 推荐系统的个性化机制为 LLM 及其智能体的发展注入了新动力，涵盖个性化 LLMs（personalized LLMs）、人格化智能体（agents with persona）以及具备个性化记忆的智能体（agents with personalized memory）。这些技术进一步挖掘智能体的特定需求，助力其提供更高效、灵活且注重隐私的服务。

可信推荐系统和智能体

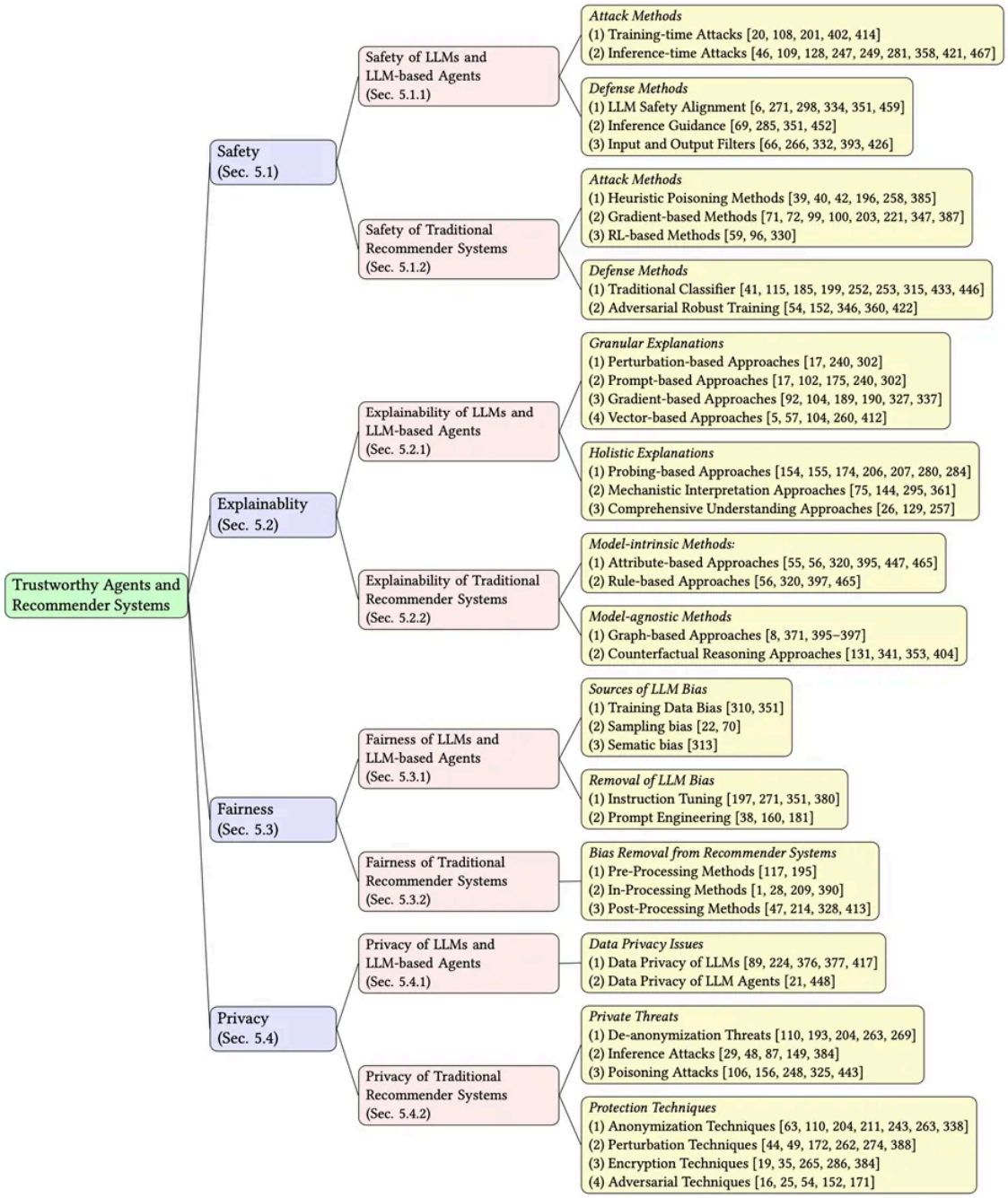

Fig. 9. Structure of Trustworthy Agents and Recommender Systems.

LLM智能体与推荐系统的可信性是现实世界应用部署中的关键挑战，主要涉及以下四个方面：

- **安全性。** LLM 及其智能体的安全性面临训练阶段攻击、推理阶段攻击以及隐私泄露等多重挑战。当前研究重点包括安全对齐、推理引导和输入/输出过滤等防御机制。推荐系统则需应对中毒攻击和去匿名化威胁，其主要防御方法包括基于分类器的异常检测和对抗性训练。推荐系统和智能体的结合还面临独特的安全挑战，如后门触发和平台负载压力，需要优化安全评估方法、强化隐私保护，并探索多智能体协作中的安全机制，为构建更安全、可信的 AI 系统提供有力支持。

- **可解释性。** LLM 及其智能体的可解释性研究包括细粒度和整体视角。细粒度视角通过特征归因方法分析输入对输出的影响，整体视角则探测或解释模型的内部机制。在推荐系统中，可解释性分为模型固有方法和模型无关方法：前者通过用户行为分析和神经符号推理直接生成解释，后者则利用反事实推理等方法生成个性化解释。评估方法包括离线测试和在线用户反馈，但后者成本高昂。未

来研究可结合检索增强生成（RAG）与知识图谱，开发涵盖记忆机制的全面框架，以提升解释可信度，推动更透明和高效的 AI 系统发展。

- **公平性。** 算法公平性对减少社会偏见、促进社会公平具有重要意义。LLM 面临来自训练数据、采样和语义编码的偏见问题，可通过指令微调和提示工程缓解不公平输出。推荐系统需应对选择偏见、曝光偏见和流行度偏见等问题，可通过数据预处理、训练过程调整和结果后处理在全流程进行改进。未来研究应关注用户特定数据的学习机制、弱势用户保护，以及公平性设计的全生命周期集成，推动技术向更具包容性、透明性和社会信任的方向发展。

- **隐私性。** LLM 及其智能体以及推荐系统的隐私问题涵盖多个维度，包括用户隐私和平台隐私两个方面。隐私威胁主要表现为去匿名化、推理攻击和中毒攻击，可能导致用户身份暴露、敏感信息泄露或系统完整性受损。隐私保护技术包括匿名化（如 k-匿名）、扰动（如差分隐私）、加密（如同态加密）和对抗性方法，通过优化数据处理和架构设计，提升系统的隐私安全性。未来研究应聚焦于数据清理、隐私过滤、分布式架构优化和用户隐私控制，并结合个性化与公平性，以构建安全可信的生态系统，平衡隐私保护与功能性能。

未来方向、挑战与机遇

我们继续从推荐系统和智能体互利共生的视角探讨新兴趋势、未来研究方向与机遇：一方面，LLM 智能体如何提升推荐系统；另一方面，推荐系统如何反过来增强 LLM 智能体的能力。

将 LLM 智能体引入推荐系统是一次革命性的发展，但其发展仍面临以下关键挑战：

- **多智能体协作处理复杂任务。** 通过将不同智能体分配到特定子任务，显著提升复杂用户查询的处理能力。
- **增强用户交互。** 开发主动互动型智能体，动态预测用户需求，实现更自然的人机对话和个性化推荐。
- **优化记忆与知识表示。** 采用短期与长期记忆的分段管理机制，并结合反思式记忆更新，提升智能体的上下文感知能力。
- **可扩展性与高效性。** 利用并行处理、高效算法和云架构，在大规模数据检索中降低延迟，增强系统的适应性。
- **伦理考量。** 设计透明、公平的推荐机制，确保高风险领域（如金融、医疗）中智能体决策的可信性与合规性。

相应地，推荐系统在优化LLM智能体性能方面具有重大潜力，并为未来研究提供方向：

- **工具与记忆推荐。** 动态推荐工具、API或外部知识资源以优化任务执行；筛选相关记忆片段，高效处理复杂的用户历史数据。
- **个性化智能体。** 为用户提供领域定制化建议（如代码辅助、客户服务、健康管理），以提升智能体的适用性与实用性。
- **规划推荐。** 更自适应地将复杂任务分解为简单步骤，提升推理效率并减少决策错误。
- **信任与可解释性。** 通过生成直观解释增强智能体决策透明性，构建可信且可解释的AI框架，以提升用户信任。

总结

由LLM驱动的智能体与推荐系统之间形成了互利共生的关系，其未来既充满机遇，也面临挑战。LLM智能体有望彻底革新用户与推荐系统的交互方式，将传统的被动推荐引擎升级为动态、交互式且适应性强的系统。同时，推荐系统通过指导工具使用、优化记忆检索和提供结构化规划，可进一步增强智能体的功能与性能。解决关键技术挑战将推动可扩展、可信赖的智能系统的发展。随着研究的深入，二者的有机结合有望实现高度个性化与主动响应的解决方案，从而大幅提升用户体验。