LLMGreenRec：面向可持续电子商务的基于大语言模型的多智能体推荐系统

Hao N. Nguyen*、Hieu M. Nguyen†‡、Son Van Nguyen*、Nguyen Thi Hanh†
*计算机科学学院，PHENIKAA大学，河内，越南；
†跨学科数字技术学院，PHENIKAA大学，河内，越南；
‡通讯作者，PHENIKAA大学，河内，越南。
*22010115@st.phenikaa-uni.edu.vn，†‡hieu.nguyenminh1@phenikaa-uni.edu.vn，
*son.nguyenvan@phenikaa-uni.edu.vn，†hanh.nguyenthi@phenikaa-uni.edu.vn

arXiv:2603.11025v1 [cs.MA] 11 Mar 2026

摘要——**电子商务中日益增强的环境意识，要求推荐系统不仅引导用户选择可持续产品，还要尽量减少其自身的数字碳足迹。传统的基于会话的系统以短期转化为优化目标，往往无法捕捉用户对环保选择的细微用户意图，从而使绿色意向与行动之间的差距持续存在。为解决这一问题，我们提出LLMGreenRec [1]，，一种利用大语言模型（大语言模型）促进可持续消费的新型多智能体框架。通过对用户交互的协同分析以及迭代式提示词优化，LLMGreenRec的专用智能体推断以绿色为导向的用户意图，并优先给出环保产品推荐。值得注意的是，这种由意图驱动的方法还能减少不必要的交互与能源消耗。在基准数据集上的大量实验验证了LLMGreenRec在推荐可持续产品方面的有效性，展示出一种促进负责任数字经济的稳健解决方案。**

索引词—**基于会话的推荐系统、大语言模型、多智能体、可持续性、电子商务**

一、引言

在电子商务蓬勃发展的背景下，用户的每一次点击和搜索查询不仅是消费者行为，也会消耗能源资源，为数据中心供能。当全球每天数十亿次交互累积起来时，在线购物的碳足迹便成为一个值得关注的问题。因此，我们面临双重挑战：引导用户选择环保产品，同时优化交互流程本身，以尽可能减少能源浪费。

证据表明，消费者对环境问题的认知正在明显增强，对可持续产品的需求也在不断增长。根据普华永道《2024年消费者之声调查》报告，54%的越南消费者愿意为采用回收材料或可持续材料制成的产品多支付至多10%的费用 [1]。在国际上，德勤《2024年全球Z世代与千禧一代调查》指出，每10名Z世代与千禧一代中有2人曾更换工作，以使工作与其环境价值观更加契合，这两个群体也愿意为可持续产品支付更高价格[2]。然而，一个显著的矛盾依然存在：大多数顾客所表达的可持续消费意愿与其实际购买行为并不一致。波士顿咨询2022年针对八个国家19,000名消费者开展的调查表明，尽管高达80%的消费者声称在日常购买中会考虑可持续性，但只有1%–7%的人会据此改变购买习惯 [3]。

当一位满怀善意的用户开始在线购物时，这种言行差距最为明显。他们的良好意愿往往受电子商务平台复杂结构的阻碍：平台上信息量庞大，传统产品占据主导地位。这使得寻找和评估环保替代品变得困难，导致决策疲劳，最终让用户转而选择更熟悉、更方便的选项。在绿色经济中，这不仅是消费者面临的问题，也给电子商务企业带来了技术挑战——用户的每一次操作，无论是滚动浏览还是搜索，都会消耗服务器能耗。单次搜索或滚动操作看似微不足道，但累积起来却会大幅增加数字碳足迹。

为了优化搜索流程、减少冗余操作并缩短用户旅程，人们开发并部署了推荐系统。然而，以绿色经济的目标为衡量标准，传统系统存在固有局限。具体而言，这些系统根据热门度数据来预测最有可能发生的下一步行动，却无意中造成了不公平的竞争环境。可持续产品的供应商通常是资源有限的小型企业，难以在由大型品牌主导的数字市场中争夺关注。这些现有推荐系统经常忽视小众产品，使其几乎无法被目标潜在客户发现。

这一问题的根源在于当前基于会话的推荐系统存在不足。它们主要旨在最大化短期转化率，缺乏识别更深层用户意图的能力，例如用户希望做出负责任选择的愿望。因此，迫切需要新一代推荐系统：不仅能够智能地捕捉用户意图，还足够灵活，能够将环境标准纳入其推荐流程。开发此类系统不仅是技术上的改进，更是打破现有壁垒、促进绿色产品市场发展以及实现可持续且负责任的数字经济的战略性关键步骤。

为满足这一需求，本文聚焦于应用一种

[1]我们的源代码和数据可在此处获取：https://github.com/haongocng/LLMsGreenRec

先进方法：构建高性能、基于会话的推荐系统，以识别用户意图并优先推荐绿色产品。借助大语言模型（LLM）的复杂推理能力，该系统能够以更少的交互准确预测用户需求，从而缩短搜索时间并降低运行能耗。同时，系统将利用对用户意图的理解，主动推荐环保替代品。最终目标是打造一种既智能便捷，又节能环保的购物流程，为绿色经济的发展作出切实贡献。

II. 相关工作

基于意图的会话推荐并不是推荐系统领域中的全新主题。从早期的基于会话的推荐系统研究开始，研究人员逐渐转向利用用户意图来提高推荐准确率。随着深度学习以及近年来大语言模型的发展，该领域取得了显著进展，尤其是在融入可持续性标准以推荐绿色产品和促进环保消费方面。以下各节将回顾相关研究，包括传统的基于会话的推荐系统、基于意图的会话推荐、大语言模型的应用、多智能体系统，以及绿色产品推荐研究。

A. 基于会话的推荐系统

基于会话的推荐系统专注于分析用户在特定会话中的短期行为序列，通常不依赖长期历史数据或用户画像。1999年，Schafer 等人 [4] 提出利用交互历史和个人偏好来推荐产品，旨在提高亚马逊和易贝等电子商务平台上的销售额。该组作者 [5] 随后开发了新方法，利用产品之间的关系和用户行为来改进对客户需求的预测。

早期推荐系统面临的主要挑战之一是海量数据。2000年，Sarwar等人 [6]提出了一种奇异值分解（奇异值分解）算法，以降低数据维度，从而减少计算时间，并更有效地生成潜在产品列表。到2002年，同一研究组 [7] 引入了一种基于最近邻算法并结合协同过滤的客户聚类方法，以提高准确性。Huang等人 [8] 后来通过使用多关系图改进了这一方法，其中产品与用户之间的链接由权重表示，在与其他推荐技术结合时表现出更优的性能。

深度学习的出现标志着基于会话的推荐方法发生了重大范式转变。2016年，Hidasi 等人 [9]首次将分层循环神经网络（RNNs）应用于挖掘短期会话数据中的潜在特征。在这一成功之后，Tan 等人[10] 对 RNNs 进行了优化，使其能够更快速地处理会话数据，并取得了显著成效。2017年，Li 等人 [11] 在 RNNs 中引入注意力机制，使模型能够关注会话中最重要的项目。Wu 等人 [12]通过引入 SR-GNN 增强了基于会话的推荐系统。该模型利用图神经网络（GNNs）将会话表示为图，从而更好地捕捉项目之间的关系。近年来，研究人员注意到，RNNs 并不适用于随机性较高的会话——Gwadabe 等人 [13] 提出了基于 GNN 的模型 GRASER，以克服这一局限。

B. 基于意图的会话推荐系统

基于意图的会话推荐通过聚焦于识别每个会话中的用户意图（例如与产品交互或搜索信息）来扩展基于会话的问题。该领域的方法可分为两大类：假设单一用户意图的方法，以及旨在学习多种意图的方法。

对于单意图假设，STAMP [14] 强调用户的最后一次行为是主要意图信号。MSGAT [15]在此基础上进行了改进，纳入了相似会话中的局部和全局信息。然而，这种方法未考虑到单个会话可能包含多个意图，因此可能会忽略重要信息。

为解决这一问题，后续研究着重于识别同一会话中的多个意图。NirGNN [16]使用注意力机制来学习多个意图。MCPRN 方法 [17] 应用路由通道对每件物品的意图进行分类。IDSR [18] 将物品表示投射到多个意图空间中，而 HIDE [19] 则将物品嵌入划分为代表不同意图的片段。此外，MIHSG [20] 和 STAGE [21] 在不同粒度层级上学习意图。尽管这些方法较为先进，但它们往往假定每个会话的意图数量固定，且其推荐过程缺乏透明度。

C. 可持续推荐系统

可持续推荐系统旨在根据与环境影响和绿色消费行为相关的数据，鼓励用户选择环保产品。该领域的早期研究主要采用协同过滤和机器学习技术，根据碳排放和再生材料含量等绿色指标对产品进行分类。一项关于可持续性推荐系统的综合综述指出，这些系统不仅能降低能源消耗，还能通过优先推荐获得绿色认证的产品，助力实现联合国可持续发展目标（SDGs） [22]。

人工智能（AI）的集成提升了绿色产品推荐的有效性。Dinesh 等人 [23] 提出了一种人工智能驱动的环保时尚推荐系统，用于分析用户行为，并推荐采用再生材料制成的服装，从而帮助降低时尚产业的环境影响。同样，一种混合推荐系统将产品功能数据与其组成部分的可持续性档案相结合，用于向消费品制造业推荐绿色替代材料，已被开发—

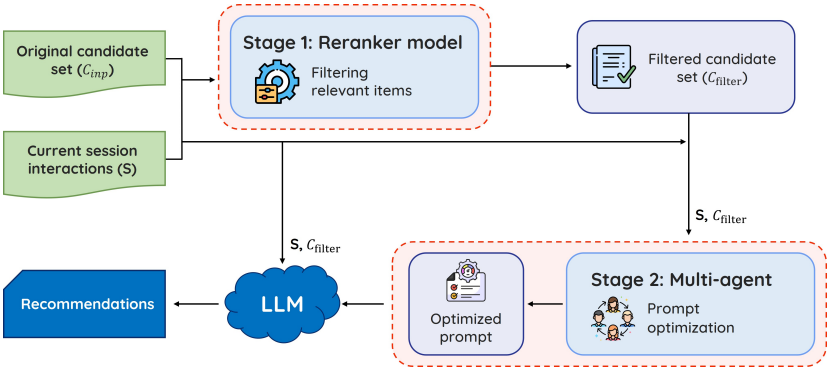

图 1：LLMGreenRec 的整体工作流程

用于向消费品制造业推荐绿色替代材料 [24]。然而，缺乏标准化的绿色指数数据以及情境的复杂性仍是主要限制，这凸显了采用大语言模型等先进模型以提升准确性和个性化程度的必要性。

D. 大语言模型（*LLM*）的应用

GPT 和 LLaMA 等大语言模型的出现，凭借其卓越的语义理解和上下文理解能力，为基于会话的推荐系统和基于意图的推荐系统开辟了新的途径。零样本下一项推荐研究 [25]是最早尝试使用零样本提示将大语言模型应用于基于会话的推荐系统、且无需对模型进行微调的研究之一。Hou 等人 [26] 利用大语言模型的上下文学习能力，根据会话中物品的顺序和新近程度，将其用作排序工具。此外，BIGRec [27] 和 GPT4Rec [28]等模型经过微调，更适用于推荐领域，并提升了系统性能。同时，RecInterpreter[29] 探索了如何解码大语言模型的表示，以将其应用于序列推荐任务。Sun 等人提出的 PO4ISR 研究 [30] 通过自动化提示词优化，利用 ChatGPT 改进基于意图的会话推荐系统，标志着一项重大进展。在 PO4ISR 之后，MACRec 技术 [31]应运而生，进一步推动了该领域的发展，重点关注在推荐系统中使用多个协作智能体。MACRec 利用多个大语言模型之间的协同机制，每个模型都可以承担不同的角色，从而提升整体性能。

近年来，大语言模型已被用于增强可持续推荐系统，利用其自然语言处理能力分析用户的绿色意向和环保产品数据。某环保产品推荐系统采用 Llama-2 作为核心引擎，根据自然语言查询生成个性化嵌入，并依据再生材料和低碳足迹等绿色因素推荐产品，从而提升用户满意度并促进可持续消费 [32]。同样，Zhou 等人展示了大语言模型能够通过优化推荐来降低能源消耗，最终助力实现全球环境目标 [33]。在一个更具体的领域中，一款面向产消者的推荐系统使用 Scikit-llm 和零样本分类器评估可持续性场景，例如基于实时数据推荐可再生能源产品 [34]。此外，大语言模型还被集成到可持续旅游系统中，利用检索增强生成（RAG）提升语境相关性并降低环境影响 [35]。这些研究凸显了大语言模型处理复杂可持续性数据的潜力。

III. 方法论

本节介绍一种名为 LLMGreen-Rec 的新型架构，旨在推荐高质量且相关的绿色产品。如图1所示，系统有两个关键输入：原始候选集（C_{inp}）和当前会话交互（S）。这些输入经过两个阶段处理：**重排序模型**和**多智能体**。第一阶段根据会话交互过滤掉不相关物品，从而缩小 C_{inp} 的规模，并得到更聚焦的候选集（C_{filter}）。第二阶段采用多智能体系统，由专用智能体分别评估用户相关性和可持续性属性。最后，大语言模型综合这些输出进行推理，选择兼顾两项标准的最佳产品集合。

A. 使用重排序模型过滤相关物品

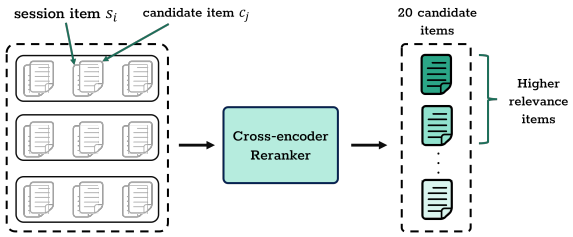

图2：交叉编码器重排序器的整体流程

整体架构的第一阶段是交叉编码器重排序器 [36]，，它作为相关性过滤步骤，如图2所示。该模型处理两个关键输入：按时间顺序排列的当前会话交互 $S = [S_1, S_2, \ldots, S_n]$，其中每个 s_i 表示用户交互过的商品，以及初始候选集，

$C_{\text{inp}} = [c_1, c_2, \ldots, c_{100}]$。它由从整个商品目录中随机选择100个产品构成，并确保始终包含目标商品。交叉编码器通过系统性地评估 C_{inp}中每个候选商品相对于用户会话交互 S的相关性来运作。首先，过滤过程从当前会话 S和初始候选集 C_{inp}中提取商品开始。随后，对于每个商品 $s \in S$，交叉编码器将其与每个候选商品$c \in \mathcal{C}_{\text{inp}}$ 配对，形成一对（s, c）。由句子转换器和分类器组成的预训练交叉编码器模型处理每一对。句子转换器生成联合嵌入，以捕捉 s与 c之间的语义关系，分类器则输出相关性分数 $\text{score}^{\text{pair}}$（$s, c$），范围从0到1。对于每个候选商品 c，总体相关性分数计算为所有会话商品上成对分数的平均值：

$$\text{score}(c) = \frac{1}{|S|} \sum_{s \in S} \text{score}_{\text{pair}}(s, c).$$

这种平均方法确保候选商品的相关性能够反映其与整个会话上下文的契合程度。最后，根据平均相关性分数对 C_{inp}中的候选商品进行排序，并选取得分最高的前 20 件商品，组成过滤后的候选集 $\mathcal{C}_{\text{filter}}$。此步骤缩小了候选池，同时保留了最相关的商品。该阶段的输出是过滤后的候选集 $C_{\text{filter}} = [c'_1, c'_2, \ldots, c'_{20}]$，其中包含在会话 S 中被认为最相关的前 20 件商品。将候选商品数量从 100 件减少到 20 件，可确保后续阶段能够将计算资源集中用于高质量的候选子集。

B. 基于多智能体框架的提示优化

为应对满足用户需求与促进绿色消费这一双重挑战，第二阶段采用了如图1所示的多智能体大语言模型（LLM）框架。该系统旨在理解并预测用户的真实意图，优化推荐过程，并特别关注绿色产品。系统的核心是自动化提示词优化过程，该过程充分利用大语言模型的卓越推理能力。最终目标是打造一个不仅准确度高，还能主动、自然地推荐可持续替代品的推荐系统，帮助弥合绿色消费意愿与实际用户行为之间的差距。

作为输入，系统会获取用户在当前会话中互动过的产品信息，以及从阶段 1 筛选出的 20 个产品组成的候选列表。这组经过精炼的候选项作为起点，使大语言模型智能体能够专注于最相关的选项，而非庞大且未经优化的数据集。

基于这个缩减后的候选集，我们设计了一套由六个基于大语言模型的智能体构成的流程，用于分析会话交互的语义，并预测用户感兴趣的下一个物品。如图3所示，这六个智能体分别是**评估、DetectError、原因推断、RefinePrompt、增强** 和 **选择**。每个智能体都有特定的功能和任务，它们密切协作，并根据预测过程的反馈持续改进提示，以提升系统的整体性能。

1) 评估智能体

评估智能体的输入包括初始提示词、会话数据和精炼候选列表。其输出是产品的排序列表，通常以文本字符串或 JSON对象的形式呈现，以便在后续步骤中轻松处理。该智能体作为关键的初步测试，帮助系统评估并调整初始提示词，以实现个性化优化。评估智能体运用大语言模型先进的自然语言处理和语义推理能力，使其能够理解用户交互历史中的复杂情境。例如，如果用户查看了与某个特定领域相关的产品，大语言模型便可推断其意图，并优先推荐类似商品。与传统方法（例如协同过滤）相比，大语言模型具有更高的灵活性和更准确的预测能力，尤其适用于稀疏的长期历史数据。此外，该智能体还利用会话数据为每位用户个性化生成排序列表，侧重于近期行为而非普遍趋势，从而提高推荐的相关性。该智能体生成的排序列表是其他智能体（如 DetectError 和 RefinePrompt）的关键输入，形成流畅的提示优化工作流。评估智能体是迭代系统的一部分，后续智能体的反馈会用于改进提示，形成正反馈回路，使系统随着时间推移逐步提高准确性。

2）错误检测智能体

随后，评估智能体生成的排序列表会传递给DetectError智能体进行错误检查。如果真实目标商品的排名低于某个阈值（例如，未进入候选列表前10名），该智能体就会将其标记为错误案例。该智能体的输入是预测的排序列表以及目标商品的实际排名位置；输出则是一个简单标记，用于指示该会话是否存在错误。该智能体的优势在于能够敏锐地发现偏差，使系统迅速响应无效提示，并维持较高的推荐质量。在基于大语言模型的系统中，提示对于模型如何解读数据至关重要，但提示未必总是最优的。如果缺少检测此类问题的机制，系统可能会继续使用欠佳的提示，导致建议不准确，并损害用户体验。该智能体充当质量控制过滤器，能够迅速识别表现不佳的提示，并通知其他智能体及时进行调整。

3) InferReason 智能体

检测到错误后，InferReason 智能体会分析其根本原因。它的输入包括错误检测智能体提供的错误报告、当前提示以及交互历史。随后，它会推断可能的原因，例如“提示的上下文针对性不足”或“提示忽略了用户近期的行为”。该智能体通过大语言模型的

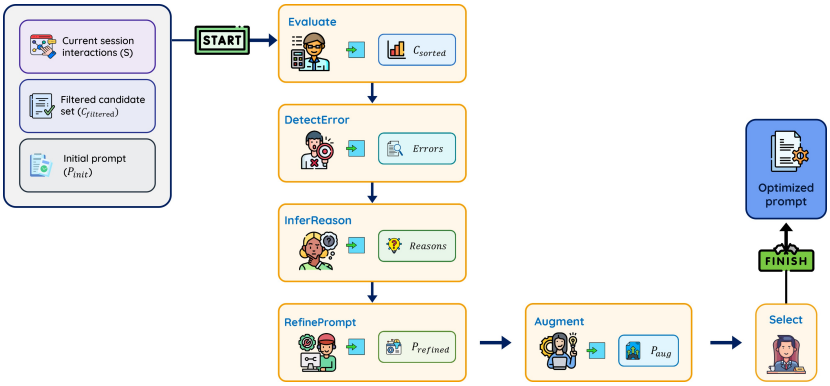

图3：多智能体系统的整体流程

自我反思机制生成一系列逻辑假设，用以解释错误。例如，收到错误后，智能体可能会思考：“为什么目标商品没有进入前 10 名？”，然后利用大语言模型分析输入数据的语义，得出结论：“提示没有强调用户对高科技产品的偏好”。这种推理不仅有助于系统纠正错误，也能让系统从错误中学习，为提示改进提供可靠依据。该智能体提供详细的推断，帮助系统有效调整提示，使系统无需持续人工干预也能变得更加智能、适应性更强。从长远来看，该智能体通过有针对性的逻辑假设优化提示优化过程，与随机试错相比，可节省运行时间并提升排序质量。

4) RefinePrompt 智能体

根据 InferReason 智能体得出的原因，RefinePrompt 智能体负责修改提示，以解决已发现的问题。其输入是旧提示和原因列表，输出是经过改进的新提示。该智能体会调整措辞或添加上下文信息，例如产品类别或会话焦点。改进后的提示会封装在 <START> 和 <END> 标签之间，以便轻松提取。该智能体的灵活性有助于创建更具体的提示，使其准确反映会话的意图和上下文，从而提升后续优化循环中的排序表现。该智能体至关重要，因为它会基于可靠的分析主动改进提示，确保调整针对问题的根本原因。这使提示更有效，也更能适应具体情况，摆脱一刀切的方法。

5) Augment 智能体

为增强对提示变体的探索，Augment 智能体会生成经过优化的提示的变体。该智能体以 RefinePrompt 智能体提供的提示为输入，生成 3–5 个不同版本，在保留核心含义的同时改变表达方式。例如，如果原始提示是“根据交互历史中的技术兴趣对产品进行排序”，Augment 智能体可能会生成“根据近期与科技相关的购物意向对产品进行排名”。这是通过利用大语言模型的释义能力实现的。该智能体引入语言多样性，并为特定用例中的最佳措辞扩展搜索空间。通过主动扩展搜索空间，Augment 智能体可避免系统拘泥于单一表达方式，从而使其能够适应多样且复杂的场景。

6) Select 智能体

Augment 智能体生成提示变体后，Select 智能体会选出最佳方案。它采用置信上界（上置信界） [37] 算法来平衡两个目标：

- 利用：选择表现良好的提示。
- 探索：测试新的提示，看看它们是否更好。

智能体按如下方式计算每个提示的置信上界值：

$$\text{UCB} = \frac{\text{reward}}{\text{selection}} + \text{explore_param.} \sqrt{\frac{\log(t)}{\text{selection}}}$$

奖励通过评估用户感兴趣的物品是否排名靠前来反映提示的质量；选择表示该提示被测试的次数；t 表示所有提示进行的试验总次数；探索参数则用于调整对新提示的探索程度。

置信上界值最高的提示将被选用于下一次迭代，具体如算法10所示。这种方法可以防止系统陷入单一提示，并通过探索新选项鼓励持续改进，从而随着时间推移增强其预测能力。

IV. 实验与结果

本文探讨两个主要研究问题：

- **研究问题1：** LLMGreenRec的多智能体模块能否在一般推荐任务中优于基线模型?
- **研究问题2：** LLMGreenRec向用户推荐相关且可持续产品的效果如何?

A. 实验设置

1) 数据集

我们在三个真实世界数据集上评估 LLMGreenRec 的性能。第一个是 MovieLen-1M (ML-1M) [38], ，它提供电影的用户评分。另外两个数据集来自

算法 1置信上界

$P = \{p_1, p_2, \ldots\}$, $R[p] = 0$, $S[p] = 0$ 对所有 $p \in P$, $T = 0$
对于 $i = 1$ 到最大试验次数 **执行**
  $T \leftarrow T + 1$
  对 每个 $p \in P$ **执行**
    $$\text{置信上界}[p] = \begin{cases} +\infty & \text{如果 } S[p] = 0 \\ \frac{R[p]}{S[p]} + \sqrt{\frac{2\ln(T)}{S[p]}} & \text{否则} \end{cases}$$
  结束循环
  sel $= \arg\max_p$ 置信上界$[p]$
  $R[\text{sel}] \leftarrow R[\text{sel}] + $ 评估(sel)
  $S[\text{sel}] \leftarrow S[\text{sel}] + 1$
结束 for

亚马逊：Games [39] 数据集，包含电子游戏评分；以及 Bundle 数据集 [40], ，提供电子、服装和食品类别的基于会话的数据，并带有显式意图标注。我们的基于会话的数据集中的每个实例都由两个部分构成：当前会话交互，即用户按时间顺序接触过的产品序列；以及目标商品，即用户最终选择的产品。我们为每个会话随机抽取 100 个商品，以构成候选集，作为筛选和提示优化阶段的初始集合。

2) 基线方法

LLMGreenRec 与一套全面的基线方法进行了基准比较，这些基线分为三大类。第一类为传统方法，例如用于推荐热门商品的 MostPop[41] 、用于会话级相似度的 SKNN [42], 以及基于马尔可夫链的矩阵分解方法 FPMC [43], 。第二类为基于深度学习的方法，分为单意图模型（如基于 RNN 的 NARM [11],、聚焦最后一项的 STAMP [14], 以及基于图的 GCE-GNN [44]）和多意图模型（如 MCPRN [17], 、HIDE [19],和 Atten-Mixer [45], ），后者旨在捕捉更多样化的用户意图。最后一类为其他基于大语言模型的方法，包括零样本提示器 NIR [25] 以及近期表现优异的 PO4ISR [30] 模型。

3) 评估指标

遵循意图驱动的会话推荐领域的最新实践 [30], ，本研究采用两项评估指标：命中率（HR@K）和归一化折损累计增益（NDCG@K），其中 K 设为 1 或 5，分别用于评估排名第一的项目和排名前五的项目。对于这两项指标，数值越高表示排序结果越理想。为确保比较公平，所有基于大语言模型的方法均基于同一个基础模型 Meta-Llama-3 [46] 实现。

B. 实验结果

1）总体比较（研究问题*1*）

为回答研究问题1，表1展示了通用推荐任务的性能对比。该评估剥离了 LLMGreenRec 的多智能体模块，使其在不优先考虑可持续性的情况下，根据随机选取的 20 项候选集进行推荐。实验结果得出了若干重要发现。首先，所提出的 LLMGreenRec 模块在所有数据集和评估指标上，始终显著优于其他所有基线方法。即使与 PO4ISR 和 NIR 等性能优异的基于大语言模型的方法相比，这种优势也尤为明显。相较于强劲的 PO4ISR 基线，该模块在 HR@5（Games数据集）上的相对提升最高达 26.6%，在 NDCG@5（Bundle 数据集）上的相对提升最高达 40.7%。如此显著的性能差距验证了所提出的多智能体优化设计的有效性，也凸显了捕捉复杂且动态变化的用户意图的必要性，而这正是 LLMGreenRec 模块的优势所在。

2）可持续推荐性能（研究问题*2*）

为回答研究问题2，表2考察了完整的两阶段 LLMGreenRec 架构，评估其在用户会话中引入相关可持续替代品的效果。在 Bundle 数据集上进行评估时，LLMGreenRec 表现出色，HR@1 达到 0.3950，HR@5 达到 0.5504。这些分数表明，在近 40% 的情况下，正确的可持续产品成功排在首位；而在超过 55% 的情况下，它出现在排名前五的推荐中，充分彰显了该模型的卓越效能。这些发现为解决可持续电子商务面临的双重挑战作出了重要贡献：推广绿色产品，同时最大限度地减少平台的能源足迹。通过准确捕捉用户意图，该系统不仅善于引入相关的可持续替代品，还能简化用户的发现流程。这使购物过程更短、更高效，减少了交互次数，从而直接降低相关的数据中心能耗，并应对在线零售的碳足迹问题。

V. 结论

受大语言模型先进推理能力的启发，本研究提出了两阶段 LLMGreenRec 框架，用于在用户会话中推荐可持续产品。该框架首先进入候选项筛选阶段，由重排序模型筛选并排列潜在商品，从而生成规模更小、相关性更高的候选集。框架的核心是一个协作式多智能体系统，其中六个专用智能体在闭环反馈周期中协同工作。该系统的独特之处在于，它能够自动识别可持续性推荐中的失误并推断其根本原因，随后迭代优化提示，为大语言模型选出效果最佳的提示。这种自我纠正机制能够更精确、灵活地捕捉用户行为和意图。在基准数据集上开展的大量实验表明，与现有方法相比，LLMGreenRec 能够提供更高质量的可持续推荐。我们认为，这项研究为价值驱动型电子商务提供了新范式，助力用户做出更明智、更可持续的选择，并为构建更复杂、符合伦理规范的个性化系统铺平道路。

表1： 推荐方法比较　　在三个数据集上的表现，最佳结果以粗体标出，“–”表示数值极小　　s

数据集	指标	传统方法		单意图方法			多意图方法			基于大语言模型的方法	
		MostPop	SKNN	FPMC	NARM	STAMP	GCE-GNN	MCPRN	HIDE	Atten-Mixer	NIR	PO4ISR	**LLMGreenRec**
ML-1M	HR@1	0.0004	0.1270	0.1132	0.1692	0.1584	0.1312	0.1434	0.1498	0.1490	0.0572	0.2000	**0.2550**
	HR@5	0.0070	0.3600	0.3748	0.5230	0.5078	0.4748	0.4788	0.4998	0.4932	0.2326	0.5510	**0.6110**
	NDCG@1	0.0004	0.1270	0.1132	0.1692	0.1584	0.1312	0.1434	0.1498	0.1490	0.0572	0.2000	**0.2550**
	NDCG@5	0.0053	0.2530	0.2464	0.3501	0.3367	0.3044	0.3157	0.3256	0.3216	0.1436	0.3810	**0.4391**
游戏	HR@1	–	0.0020	0.0498	0.0572	0.0556	0.0692	0.0522	0.0696	0.0530	0.1168	0.2588	**0.4580**
	HR@5	–	0.0020	0.2564	0.2574	0.2586	0.2744	0.2416	0.2694	0.2472	0.3406	0.5866	**0.7430**
	NDCG@1	–	0.0020	0.0498	0.0572	0.0556	0.0692	0.0522	0.0696	0.0530	0.1168	0.2588	**0.4580**
	NDCG@5	–	0.0020	0.1508	0.1534	0.1555	0.1701	0.1432	0.1662	0.1475	0.2310	0.4313	**0.6143**
捆绑	HR@1	–	–	0.0398	0.0322	0.0365	0.0360	0.0360	0.0458	0.0525	0.0975	0.1697	**0.2815**
	HR@5	0.0042	–	0.2475	0.2322	0.2352	0.2237	0.2352	0.2585	0.2644	0.2832	0.4328	**0.5588**
	NDCG@1	–	–	0.0398	0.0322	0.0365	0.0360	0.0360	0.0458	0.0525	0.0975	0.1697	**0.2815**
	NDCG@5	0.0021	–	0.1395	0.1303	0.1339	0.1267	0.1490	0.1495	0.1549	0.1939	0.3040	**0.4279**

表2： 捆绑上的可持续推荐性能

数据集	指标			
	HR@1	*HR@5*	*NDCG@1*	*NDCG@5*
捆绑	0.3950	0.5504	0.3950	0.4715

参考文献

[1]PwC 越南，消费者业务首席合伙人 Nguyen Luong Hien，《响应、反思、重塑：增强消费者信任——越南消费者之声调查 2024年：亚太地区》，PwC 越南，技术报告，2024年8月。 [在线]。获取地址：https: // www.pwc.com/vn/en/publications/2024/voice-of-consumer-2024.pdf[2]德勤有限公司，《在变革的世界中有目标地生活与工作：2024年 Z 世代与千禧一代调查》，德勤，技术报告，2024年6月。[在线]。获取地址：https://www.deloitte.com/content/dam/assets-shared/ docs/campaigns/ 2024 2024 =6 /deloitte- - [3] genz-millennial-survey.pdf?dlva K. Sanghi、A. Bharadwaj、L. Taylor、L. Turquier 和 I. Zaveri，《消费者是推动绿色消费主流化的关键》，波士顿咨询集团，2022年9月。 [在线]。获取地址：https://www.bcg.com/publications/2022/ consumers-are-the-key-to-taking-sustainable-products-mainstream [4] J. B. Schafer、J. Konstan 和 J. Riedl，《电子商务中的推荐系统》，载于第 *1* 届 *ACM* 电子商务会议论文集（*EC '99*），ACM 国际会议论文集系列，美国科罗拉多州丹佛，1999年11月，第 158–166 页。 [在线]。获取地址：https://dl.acm.org/doi/pdf/ 10 1145 336992 337035 . / . J. B. Schafer、J. A. [5] Konstan 和 J. Riedl，《电子商务推荐应用》，《数据挖掘与知识发现》，第 5 卷，第 1–2 期，第 115–153 页，2001年1月。 [在线]。获取地址：https: // link.springer.com/article/ 10 1023 1009804230409 . / [6] A: B. M. Sarwar、G. Karypis、J. Konstan 和 J. T. Riedl，《在推荐系统中应用降维：案例研究》，载于第 *6* 届 *ACM SIGKDD* 知识发现与数据挖掘国际会议论文集，电子商务网络挖掘研讨会：挑战与机遇（*WEBKDD '00*），美国马萨诸塞州波士顿，2000年8月。 [在线]。获取地址：https://files.grouplens.org/papers/webKDD00.pdf [7] B. M. Sarwar、G. Karypis、J. Konstan 和 J. Riedl，《面向大规模电子商务的推荐系统：利用聚类实现可扩展的邻域构建》，载于第五届计算机与信息科学国际会议论文集，美国明尼苏达州明尼阿波利斯，2002年。[在线]。获取地址：https://files.grouplens.org/papers/sarwar cluster.pdf [8] Z. Huang、W. Chung 和 H. Chen，《电子商务推荐系统的图模型》，《美国信息学会期刊》科学与技术，第55卷，第3期，第259–274页，2004年2月。 [ ] 10 1002 10372 在线。可获取：[9] https://doi.org/ . /asi。B. Hidasi、A. Karatzoglou、L. Baltrunas和D. Tikk，“使用循环神经网络的基于会话的推荐”，收录于国际学习表征会议（*ICLR*），2016年。 [在线]。可获取：https://hidasi.eu/assets/pdf/gru4rec iclr16.pdf [10] Y. K. Tan、X. Xu和Y. Liu，“面向基于会话推荐的改进循环神经网络”，收录于第*1*届推荐系统深度学习研讨会论文集（*DLRS '16*）。美国纽约州纽约：ACM出版社，2016年，第17–22页。 [在线]。可获取：http://doi.acm.org/ 10 1145 2988450 2988452 . / 。J. Li、P. Ren、Z. Chen、Z. [11] Ren、T. Lian和J. Ma，“神经注意力会话推荐”，收录于*2017*年*ACM*信息与知识管理会议论文集（*CIKM '17*），新加坡，2017年11月，第1419–1428页。 [在线]。可获取：https://doi.org/ 10 1145 3132847 3132926 . / 。 S. Wu、Y. Tang、Y. Zhu、L. Wang、X. Xie和T. Tan，“基于图神经网络的会话推荐”，收录于第三十三届*AAAI*人工智能会议论文集（*AAAI-19*），第33卷，第01期。美国夏威夷州檀香山：AAAI出版社，2019年7月，第346– [12] 353 [ ] 10 1609 33 01 3301346 页。在线。可获取：https://doi.org/ . /aaai.v i 。T. R. Gwadabe、M. A. M. Al-hababi和Y. Liu，“通过非序列交互改进基于会话推荐系统的图神经网络”，神经计算，第468卷，第111–122页，2022年1月。 [在线]。可获取：https://www.sciencedirect.com/science/ article/pii/S 0925231221015149 [13] Q. Liu、Y. Zeng、R. Mokhosi和H. Zhang，“Stamp：用于基于会话推荐的短期注意力/记忆优先模型”，收录于第*24*届*ACM SIGKDD*知识发现与数据挖掘国际会议论文集（*KDD '18*），英国伦敦，2018年，第1831–1839页。 [在线]。可获取：https: //doi.org/ 10 1145 3219819 3219950 . / 。S. Qiao、W. Zhou、J. Wen、H. [14] Zhang和M. Gao，“面向基于会话推荐的双通道多重稀疏图注意力网络”，收录于第*32*届*ACM*信息与知识管理国际会议论文集（*CIKM '23*），英国伯明翰，2023年，第2075–2084页。 [在线]。可获取：https://doi.org/ 10 1145 3583780 3614791 . / 。D. Jin、L. Wang、 [15] Y. Zheng、G. Song、F. Jiang、X. Li、W. Lin和S. Pan，“用于基于会话的新物品推荐的双重意图增强图神经网络”，收录于*2023*年*ACM*网络会议论文集（*WWW '23*），美国得克萨斯州奥斯汀，2023年，第684–693页。 [在线]。可获取：https://doi.org/ 10 1145 3543507 3583526 . / 。 S. Wang、L. Hu、Y. Wang、Q. Z. Sheng、M. Orgun和L. Cao，“通过混合通道目的路由网络为下一物品推荐建模多用途会话”，收录于第*28*届国际人工智能联合会议论文集（*IJCAI-19*），S. Kraus编，中国澳门，2019年，第3771–3777页。 [16] [在线]。可获取：https://doi.org/10.24963/ijcai.2019/523 [17]

[18] W. Chen、P. Ren、F. Cai、F. Sun和M. de Rijke，“利用意图感知多样化改进端到端序列推荐”，收录于第29届ACM信息与知识管理国际会议论文集（CIKM ’20），爱尔兰线上活动，2020年，第175–184页。 [在线]。可获取：https://doi.org/10.1145/3340531. 3411897[19] Y. Li、C. Gao、H. Luo、D. Jin和Y. Li，“利用意图解耦增强超图神经网络以用于基于会话的推荐”，收录于第45届ACM信息检索研究与发展国际会议论文集（SIGIR ’22），西班牙马德里，2022年，第1997–2002页。 [在线]。可获取：https://doi.org/ 10 1145 3477495 3531794 . / 。J. Guo、Y. Yang、[20] X. Song、Y. Zhang、Y. Wang、J. Bai和Y. Zhang，“学习用于基于会话推荐的多粒度连续用户意图单元”，收录于第15届ACM网络搜索与数据挖掘国际会议论文集（WSDM ’22），美国亚利桑那州菲尼克斯，2022年，第343–352页。 [在线]。可获取：https://doi.org/ 10 1145 3488560 3498524 . / 。Y. Li、C. Gao、X. Du、H. Wei、[21] H. Luo、D. Jin和Y. Li，“基于图神经网络的时空感知的基于会话推荐”，收录于第31届ACM信息与知识管理国际会议论文集（CIKM ’22），美国佐治亚州亚特兰大，2022年，第1209–1218页。 [在线]。可获取：https://doi.org/ 10 1145 3511808 3557458 . / 。A. Felfernig、[22] M. Wundara、T. N. T. Tran、S. Polat-Erdeniz、S. Lubos、M. E. Mansi、D. Garber和V.-M. Le，“面向可持续性的推荐系统：概述与研究问题”，Frontiers in Big Data，第6卷，文章编号1284511，2023年。 [在线]。可获取：https://www.frontiersin.org/articles/10.3389/fdata.2023.1284511/full [23] T. S. Dinesh、K. Ammaiyappan、G. Janav、S. Gopikumar、R. S. Krishnan和A. Srinivasan，“面向具有生态意识消费者的新型人工智能驱动的推荐系统”，收录于2024年第8届发明系统与控制国际会议论文集（ICISC），2024年，第376– 382页。[24] L. Chizzali、A. Nesti以及SDSC团队，“面向更可持续成分替代方案的自动化推荐系统”，案例研究，瑞士数据科学中心，7月 11 2025 [ ] 。在线。可获取：https://www.datascience.ch/case-studies/leveraging-ai-to-foster-sustainable-consumer-goods [25] L. Wang和E.-P. Lim，“使用大型预训练语言模型的零样本下一项目推荐”，arXiv预印本arXiv: *2304 03153*, 。在线。可获取：https://arxiv.org/abs/ 。 2023 [ ] 2304 03153 [26] Y. Hou、J. Zhang、Z. Lin、H. Lu、R. Xie、J. McAuley和W. X. Zhao，“大语言模型是推荐系统的零样本排序器”，收录于《信息检索进展》，计算机科学讲义丛书。Springer，2024年，第364–381页。 [在线]。可获取：https://doi.org/ 10 1007 978 3 031 56060 6 24 . / - - - - K. Bao及合著者，“推荐[27] 系统中大语言模型的双步骤落地范式”，ACM推荐系统汇刊，第3卷，第4期，2025年。[28] J. Li、W. Zhang、T. Wang、G. Xiong、A. Lu和G. Medioni，“GPT4Rec：用于个性化推荐与用户兴趣解读的生成式框架”，arXiv预印本arXiv: *2304 03879*, 。在线。可获取：https://arxiv.org/abs/ 。 2023 [ ] 2304 03879 [29] Z. Yang、J. Wu、Y. Luo、J. Zhang、Y. Yuan、A. Zhang、X. Wang和X. He，“大语言模型可以解释序列推荐系统的潜在空间”，arXiv预印本arXiv:2310.20487，2023年。 [ ] 2310 20487 在线。可获取：[30] https://arxiv.org/abs/ 。Z. Sun、H. Liu、X. Qu、K. Feng、Y. Wang和Y.-S. Ong，“用于意图驱动的会话推荐的大语言模型”，收录于第47届ACM信息检索研究与发展国际会议论文集（SIGIR ’24），美国华盛顿特区，2024年，第324–334页。 [在线]。可获取：https://doi.org/ 10 1145 3626772 3657688 . / 。Z. Wang、Y. Yu、W. Zheng、[31] W. Ma和M. Zhang，“MACRec：用于推荐的多智能体协作框架”，收录于第47届ACM信息检索研究与发展国际会议论文集（SIGIR ’24），美国华盛顿特区，2024年，第2760–2764页。 [在线]。可获取：https://doi.org/ 10 1145 3626772 3657669 . / 。C. Bondgulwar、S. Jagtap、[32] S. Shelar和P. Shevatekar，“使用大语言模型Llama-2的环保产品推荐系统”，International Journal of Innovative Research in Technology (IJIRT)，第11卷，第10期，第2699–2704页，2025年。 [在线]。可访问：https://ijirt.org/Article?manuscript[33] X. Zhou、L. Zhang、H. Zhang、Y. Zhang、X. Zhang、J. Zhang和Z. Shen，“通过推荐系统推进可持续性：一项综述”，arXiv预印本arXiv:2411.07658，2024年。 [在线]。可访问：https://arxiv.org/abs/ 2411 07658 。S.-V. Oprea和A. Bˆara，“基于大语言模型的产消者推荐系统”，Sensors，第[34] 24， 11， 3530, 卷，第期，第页。2024 [ ] 10 3390 在线。可访问：https://doi.org/。/s 24113530 [35] ¨ A. Banerjee、A. Satish和W. Worndl，“利用检索增强生成提升面向可持续城市旅行的旅游推荐系统”，载于《面向可持续性与社会公益的推荐系统》，《计算机与信息科学通信》系列，A. Banerjee和W. W¨orndl编，Springer，2025年，第2470卷，第19–34页。 [在线]。可访问：https://doi.org/10.1007/ 978-3-031-87654-7 3[36] N. Reimers和I. Gurevych，“Sentence-BERT：使用孪生BERT网络的句子嵌入”，载于《2019年自然语言处理经验方法会议暨第9届国际自然语言处理联合会议论文集》（EMNLP-IJCNLP）。中国香港：计算语言学协会，2019年11月，第3982–3992页。 [在线]。可访问：https: //aclanthology.org/D 19 1410 - R. Pryzant、D. Iter、J. Li、Y. Lee、C. Zhu和M. Zeng，“使用‘梯度下降’和束搜索进行自动提示优化”，载于《2023年自然语言处理经验方法会议论文集》（EMNLP）。新加坡：计算语言学协会，2023年，第7957–7968页。 [在线]。可访问：https://aclanthology.org/2023.emnlp-main.494/ [37] [38] F. M. Harper和J. A. Konstan，“MovieLens数据集：历史与背景”，ACM交互智能系统汇刊（TiiS），第5， 4, 2015 [ ] 10 1145 2827872 卷，第期。在[39] 线。可访问：https://doi.org/。/ J. Ni、J. Li和J. McAuley，“利用远程标注评论和细粒度方面为推荐提供解释”，载于《2019年自然语言处理经验方法会议暨第9届国际自然语言处理联合会议论文集》（EMNLP-IJCNLP），2019年，第188–197页。[40] Z. Sun、J. Yang、K. Feng、H. Fang、X. Qu和Y. S. Ong，“重新审视捆绑推荐：数据集、任务、挑战与意图感知产品捆绑机遇”，载于SIGIR ’22：《第45届ACM信息检索研究与发展国际会议论文集》，2022年，第2900–2911页。 [在线]。可访问：https://doi.org/ 10 1145 3477495 3531904 。/。Y. Ji、A. Sun、J. Zhang和C. Li，“重新审视推荐系统中的流行度基线”，arXiv预印本arXiv: [41] *2005 13829*, 。。在线。可访问：2020 [ ] 2005 13829 https://arxiv.org/abs/。[42] D. Jannach和M. Ludewig，“当循环神经网络遇见基于会话的推荐中的邻域时”，载于RecSys ’17：《第十一届ACM推荐系统会议论文集》，2017年，第306–310页。 [在线]。可访问：https://doi.org/ 10 1145 3109859 3109872 。/。S. Rendle、C. Freudenthaler和L. Schmidt-Thieme，“用于下一购物篮推荐的个性化马尔可夫链分解”，载于WWW ’10：《第19届万维网国际会议论文集》，2010年，第811–820页。 [在线]。可访问：https: //doi.org/ [43] 10 1145 1772690 1772772 。/。Z. Wang、W. Wei、G. Cong、X.-L. Li、[44] X.-L. Mao和M. Qiu，“用于基于会话推荐的全局上下文增强图神经网络”，arXiv预印本arXiv:2106.05081，2021年。 [在线]。可访问：https://doi.org/ 10 48550 2106 05081 。/arXiv。。P. Zhang、J. Guo、C. Li、Y. Xie、J. Kim、Y. Zhang、X. Xie、H. Wang和S. Kim，“通过[45] Atten-Mixer网络高效利用多层次用户意图进行基于会话的推荐”，arXiv预印本arXiv:2206.12781，2022年。 [在线]。可访问：https: // doi.org/10.48550/arXiv.2206.12781[46]Llama团队、AI @ Meta，“Llama 3模型群”，arXiv预印本arXiv:2407.21783，2024年。 [在线]。可访问：https: // arxiv.org/pdf/2407.21783