

美团大模型 Agent 实践手册

美团大模型 AGENT 实践手册.....	1
前言.....	1
第一章 美团大模型 Agent 基础认知.....	2
1.1 大模型 Agent 的定义与核心能力.....	2
1.2 美团大模型 Agent 的定位与价值.....	3
1.3 美团大模型 Agent 的发展历程.....	4
第二章 美团大模型 Agent 技术架构.....	5
2.1 龙猫大模型 (LongCat-Flash-Chat) 核心架构.....	5
2.2 模型训练流程与策略.....	5
2.3 模型能力评估矩阵.....	7
第三章 美团各业务线大模型 Agent 实践.....	7
3.1 外卖业务线大模型 Agent 实践.....	8
3.2 到店业务线大模型 Agent 实践.....	10
3.3 酒旅业务线大模型 Agent 实践.....	12
3.4 共享单车业务线大模型 Agent 实践.....	14
第四章 美团大模型 Agent 开发流程.....	16
4.1 需求分析与场景定义.....	16
4.2 数据准备与预处理.....	16
4.3 模型选型与微调.....	17

4.4 Agent 架构设计与实现.....	18
4.5 测试与优化.....	19
第五章 美团大模型 Agent 工程化实践.....	20
5.1 开发工具链与平台支持.....	20
5.2 监控与运维体系.....	21
5.3 安全与合规保障.....	22
第六章 美团大模型 Agent 评估与迭代.....	23
6.1 评估指标体系.....	24
6.2 A/B 测试方法.....	25
6.3 迭代优化策略.....	25
第七章 美团大模型 Agent 最佳实践与避坑指南.....	26
7.1 跨业务线通用原则.....	26
7.2 常见问题与解决方案.....	28
7.3 成功案例分享.....	29
第八章 未来展望与技术演进.....	30
8.1 美团大模型 Agent 技术发展方向.....	30
8.2 面临的挑战与应对策略.....	31
8.3 结语.....	32
第三章 美团各业务线大模型 Agent 实践 (续)	33
3.5 生鲜零售业务线大模型 Agent 实践.....	33
第二章 美团大模型 Agent 技术架构 (续)	35
2.4 龙猫大模型架构升级 (2024)	35

2.5 模型训练效率优化.....	37
第五章 美团大模型 Agent 工程化实践 (续)	37
5.4 绿色 AI 工程实践.....	37
第六章 美团大模型 Agent 评估与迭代 (续)	38
6.4 多维度评估体系升级.....	38
6.5 智能迭代优化系统.....	39
第七章 美团大模型 Agent 最佳实践与避坑指南 (续)	40
7.4 多 Agent 协同优化案例.....	40
7.5 前沿技术落地陷阱与规避.....	41
第八章 未来展望与技术演进 (续)	42
8.4 2025-2027 技术发展路线图.....	42
8.5 生态共建与社会责任.....	43

前言

在即时零售行业蓬勃发展的当下，消费者对于商品和服务的即时性、准确性需求日益提升，这对我们美团的服务能力提出了更高的要求。在这样的行业背景下，大模型 Agent 技术应运而生，成为推动美团“零售 + 科技”战略落地的关键力量。

传统的规则引擎在面对复杂多变的业务场景时，逐渐显露出其局限性。它难以应对海量的动态数据，也无法快速适应业务模式的创新与变化。而美团自主研发的 LongCat-Flash-Chat（龙猫）大模型的出现，为我们突破这些局限带来了希望。龙猫大模型凭借其先进的技术架构和强大的智能体能力，能够更好地理解业务场景，更高效地处理复杂任务，为美团各业务线的智能化升级提供了坚实的技术支撑。

本手册旨在结合美团各业务线的实际实践，从技术架构、业务落地、工程实践等多个维度，全面阐述大模型 Agent 在美团的应用方法和经验，为各业务线的大模型 Agent 相关工作提供切实可行的指导，助力美团实现“帮大家吃得更好，生活更好”的使命。

第一章 美团大模型 Agent 基础认知

1.1 大模型 Agent 的定义与核心能力

大模型 Agent 是基于大语言模型构建的具有自主决策、自主行动能力的智能实体。它能够理解用户需求，进行任务规划，调用相关工具和资源，与环境进行交互，最终完成特定的业务任务。

在美团的业务场景中，大模型 Agent 具备以下核心能力：

- **自然语言理解能力**：能够准确理解用户通过文字、语音等方式表达的需求，包括复杂的语义、情感和意图。例如，在智能客服场景中，能够理解用户关于订单问题、退款申请等各种表述。
- **任务规划能力**：对于复杂的用户需求，能够将其分解为一系列子任务，并规划出合理的执行顺序和步骤。比如，在酒旅预订场景中，根据用户的出行时间、预算、偏好等，规划出合适的酒店、交通等预订方案。
- **工具调用能力**：能够根据任务需求，自主调用美团内部的各种工具、系统和 API，如订单系统、支付系统、推荐系统等，以获取所需的数据和资源，完成任务执行。
- **环境交互能力**：能够与用户、商户、骑手等进行实时交互，获取反馈信息，并根据反馈调整自己的行为 and 决策。例如，在骑手智能助手中，根据骑手的实时位置和路况信息，调整配送路线。
- **学习与进化能力**：通过不断地与业务场景交互和数据积累，能够持续学习和优化自身的模型和策略，提升任务处理的准确性和效率。

1.2 美团大模型 Agent 的定位与价值

在美团“零售 + 科技”的战略布局中，大模型 Agent 扮演着至关重要的角色，是实现业务智能化升级的核心引擎。

从业务定位来看，大模型 Agent 贯穿于美团的整个业务链条，涵盖外卖、到店、酒旅、共享单车等多个业务线。它不仅是连接用户、商户和平台的重要桥梁，也是提升各环节运营效率和服务质量的关键支撑。

大模型 Agent 为美团带来的价值主要体现在以下几个方面：

- **提升用户体验：**通过更精准地理解用户需求，提供个性化、高效的服务，减少用户等待时间和操作成本，增强用户满意度和忠诚度。例如，智能推荐系统为用户推荐更符合其口味和需求的商品和服务。
- **提高运营效率：**自动化处理大量重复性、规律性的业务任务，减少人工干预，降低运营成本，同时提高任务处理的速度和准确性。如智能调度系统实现订单与骑手的快速匹配和路径优化。
- **驱动业务创新：**能够挖掘业务数据中的潜在价值，发现新的业务机会和模式，为业务创新提供有力的技术支持。比如，基于用户行为数据的分析，开发新的营销活动和形式。
- **增强决策能力：**为管理层提供更全面、准确的数据分析和决策建议，帮助企业做出更科学的战略决策。例如，通过对市场趋势和用户需求的预测，调整业务布局和资源配置。

1.3 美团大模型 Agent 的发展历程

美团在大模型 Agent 领域的探索并非一蹴而就，而是经历了一个不断演进和完善的过程。

早期，美团主要采用传统的规则引擎和机器学习算法来实现部分智能化功能。在这个阶

段，我们构建了一些简单的智能系统，如基于规则的推荐系统和客服自动回复系统。这些系统在一定程度上提升了业务效率，但受限于技术架构，其智能化水平和灵活性都比较有限。

随着深度学习技术的发展，美团开始投入到深度学习模型的研发和应用中。我们尝试将卷积神经网络、循环神经网络等模型应用于图像识别、自然语言处理等领域，取得了一定的成果。例如，在图像识别方面，实现了对商品图片的自动分类和标签标注；在自然语言处理方面，提升了客服对话的理解准确性。

近年来，大语言模型技术的飞速发展，为美团大模型 Agent 的研发带来了新的机遇。美团技术团队抓住这一契机，加大了在大模型领域的投入，成功研发出了 LongCat-Flash-Chat(龙猫)大模型。龙猫大模型的推出，标志着美团在大模型 Agent 领域进入了一个新的阶段。基于龙猫大模型，我们在各业务线开展了广泛的应用实践，不断优化和完善大模型 Agent 的性能和功能。

第二章 美团大模型 Agent 技术架构

2.1 龙猫大模型 (LongCat-Flash-Chat) 核心架构

美团龙猫大模型 (LongCat-Flash-Chat) 是一款拥有 5600 亿参数的混合专家 (MoE) 语言模型，其核心架构具有两项关键创新，分别是零计算专家机制和捷径连接 MoE (ScMoE)。

零计算专家机制针对 MoE 模块中词元重要性差异，实现了动态计算资源分配。根据语境需求，模型能够激活 186 亿至 313 亿参数 / 词元，平均激活参数约为 270 亿。通过 PID 控制器调节专家偏置，将单词元平均激活参数稳定在这一范围内，从而优化了资源利用效率。这一机制使得模型在处理不同重要程度的词元时，能够合理分配计算资源，在保证模型性能的同时，降低了计算成本。

捷径连接 MoE (ScMoE) 则扩大了计算 - 通信重叠窗口，结合定制化基础设施优化，支撑了万级加速器规模的训练及高吞吐、低延迟推理。每层集成两个多头潜在注意力 (MLA) 模块和多个异构前馈网络 (FFN) 模块，并采用从第一个 MLA 输出到 MoE 模块的快捷连接，有效提升了模型的推理效率和吞吐量。

2.2 模型训练流程与策略

龙猫大模型的训练采用了多阶段流程，通过精心设计的训练策略，培育出了强大的智能

体能力。

第一阶段为预训练，采用两阶段数据融合策略聚焦推理密集型领域。在数据收集方面，整合了美团各业务线的海量数据，包括用户行为数据、商户信息数据、订单数据等。通过数据清洗、预处理和筛选，确保训练数据的质量和相关性。在预训练过程中，模型学习到了语言的基本规律和知识，为后续的训练阶段奠定了基础。

第二阶段为中期训练，主要目的是增强模型的推理与编码能力，并将上下文扩展至 128k 以适配后续需求。在这一阶段，通过引入大量的推理任务和编码任务数据集，对模型进行针对性训练。同时，采用精细化的训练策略，如调整学习率、优化器参数等，提升模型在这些任务上的性能。

第三阶段为后期训练，设计了多智能体合成框架，从信息处理、工具复杂度、用户交互三维度定义任务难度，生成需迭代推理与环境交互的复杂任务。通过这些复杂任务的训练，进一步提升模型的智能体能力，使其能够更好地应对实际业务场景中的复杂问题。

得益于架构、策略与基础设施的协同创新，LongCat-Flash 在 5600 亿参数规模下仅用 30 天完成 20 万亿词元预训练(时间可用率 98.48%)，H800 推理效率超 100 词元 / 秒 (TPS)，单百万输出词元成本 0.7 美元。

2.3 模型能力评估矩阵

为全面评估龙猫大模型的能力，我们构建了一套完善的模型能力评估矩阵，从多个维度对模型进行考量。

在通用领域能力方面，采用 ArenaHard-V2 基准进行评估，LongCat-Flash 获得 86.5 分，展现了其在通用语言理解和生成方面的强大能力。在编程能力方面，通过 TerminalBench 基准测试，获得 39.5 分，表明模型在代码生成、调试等方面具有一定的水平。

在智能体工具使用方面， r^2 -Bench 基准测试得分为 67.7 分，体现了模型在调用工具完成任务方面的能力。此外，为降低现有开源基准数据污染风险并提升评估可信度，我们精心构建了两个新基准：Meeseeks 和 VitaBench。Meeseeks 通过迭代反馈框架模拟真实人机交互以评估多轮指令跟随能力，LongCat-Flash 在该基准上的表现与前沿大语言模型相当。VitaBench 则利用真实商业场景评估模型处理复杂现实任务的能力，在此基准上 LongCat-Flash 的表现优于其他大语言模型。

通过这套评估矩阵，我们能够全面了解模型的优势和不足，为模型的优化和迭代提供有力的指导。

第三章 美团各业务线大模型 Agent 实践

3.1 外卖业务线大模型 Agent 实践

3.1.1 智能调度系统

美团外卖智能配送调度系统是世界上最规模最大、复杂度最高的多人、多点实时智能配送调度系统之一，每天匹配 50 多万外卖小哥，基于海量数据和人工智能算法，确保平均配送时长不超过 28 分钟。

在智能调度系统中，大模型 Agent 发挥着核心作用。其主要功能包括订单指派、路径规划和 ETA（预计到达时间）预测。订单指派是一个 NP-Hard 问题，需要在几十毫秒内将新生成的订单分配到最合适的骑手。大模型 Agent 通过分析骑手的实时位置、负载情况、历史配送数据等信息，结合订单的地址、商家出餐速度等因素，快速做出最优的订单分配决策。

路径规划方面，大模型 Agent 考虑了天气、路况、骑手熟练程度等多达几十个因素，为骑手规划出最优的配送路径。通过不断学习和优化，路径规划的准确性和效率持续提升，有效缩短了配送时间。

ETA 预测是智能调度系统的重要环节，大模型 Agent 基于历史配送数据、实时路况、天气等信息，精准预测订单的到达时间，并及时将信息反馈给用户和骑手，提升了用户体验和骑手的工作效率。

3.1.2 骑手智能助手

为全面提升骑手的各方面能力，美团研发团队设计了骑手智能助手，以智能耳机和语音交互为载体，通过大数据挖掘、机器学习、自然语言处理等技术，让智能助手具备复杂场景精准识别、服务智能推送，智能引导、全语音操作等能力。

大模型 Agent 在骑手智能助手中的应用，主要体现在以下几个方面：

- 复杂场景识别：能够实时识别骑手所处的场景，如商家出餐慢、交通拥堵、用户地址不明确等，并为骑手提供相应的解决方案和建议。

- 智能信息推送：根据骑手的实时位置和配送任务，及时推送相关信息，如订单新动态、路况提醒、天气预警等。
- 全语音交互：支持骑手通过语音指令完成各种操作，如接单、上报异常、导航等，减少了骑手的手动操作，提高了工作效率，同时保障了骑行安全。

通过骑手智能助手的应用，在智能、安全、便捷、精准等多个维度上，全面提升了骑手配送能力，从而提升了整个配送效率和用户体验。

3.1.3 订单异常处理 Agent

在外卖业务中，订单异常情况时有发生，如商家漏发商品、商品损坏、配送延迟等。订单异常处理 Agent 的出现，有效提高了异常订单的处理效率和用户满意度。

大模型 Agent 通过实时监控订单状态，及时发现异常订单。一旦发现异常，Agent 会自动与用户、商家和骑手进行沟通，了解具体情况。根据异常类型和严重程度，Agent 会制定相应的解决方案，如安排补发商品、退款、赔偿等。

在处理过程中，Agent 能够自主调用订单系统、支付系统等相关工具，完成解决方案的执行。同时，Agent 会记录异常处理过程和结果，为后续的数据分析和优化提供依据。通过订单异常处理 Agent 的应用，大大缩短了异常订单的处理时间，减少了用户的投诉和不满。

3.2 到店业务线大模型 Agent 实践

3.2.1 智能搜索与推荐系统

到店业务线涵盖了餐饮、休闲娱乐、丽人等多个领域，用户需求多样且复杂。智能搜索与推荐系统基于大模型 Agent 技术，为用户提供更精准、个性化的搜索结果和推荐内容。大模型 Agent 能够深入理解用户的搜索意图，包括关键词、语义、情感等。通过分析用户的历史行为数据、偏好信息以及当前的搜索上下文，Agent 能够准确把握用户的需求，为用户返回最相关的商家和服务信息。

在推荐方面，Agent 采用了多维度的推荐策略，综合考虑用户的兴趣爱好、消费能力、地理位置、口碑评价等因素，为用户推荐最合适的商家和优惠活动。同时，Agent 还能

够根据实时数据，如商家的客流量、 availability 等，动态调整推荐结果，确保推荐的准确性和时效性。

通过智能搜索与推荐系统的应用，提高了用户的搜索效率和推荐满意度，为商家带来了更多的客流和订单。

3.2.2 AI 导购与智能客服

AI 导购和智能客服是到店业务线提升服务质量和用户体验的重要手段。基于大模型 Agent 技术的 AI 导购能够为用户提供个性化的购物建议和服务推荐。

在用户浏览商家或商品信息时，AI 导购会主动与用户交互，了解用户的需求和偏好。通过分析用户的反馈和历史数据，Agent 能够为用户推荐合适的商品、套餐和服务，并解答用户的疑问。例如，在餐饮场景中，AI 导购可以根据用户的口味偏好、人数等，推荐合适的菜品和餐厅。

智能客服则能够快速响应用户的咨询和问题，提供专业、准确的解答。大模型 Agent 具备强大的自然语言理解和处理能力，能够理解用户的各种问题表述，并从知识库中获取相关信息进行回答。对于复杂的问题，Agent 会及时转接人工客服，并将相关信息同步给人工客服，提高问题解决效率。

3.2.3 商户运营助手

商户运营助手为到店商户提供了全方位的运营支持，帮助商户提升经营效率和盈利能力。大模型 Agent 在商户运营助手中发挥着重要作用，主要体现在以下几个方面：

- **数据分析与洞察**：Agent 能够对商户的经营数据进行全面分析，包括客流量、销售额、客单价、用户评价等，为商户提供详细的数据分析报告和经营洞察。商户可以根据这些信息，调整经营策略，优化产品和服务。
- **营销活动策划**：基于商户的经营状况和用户数据，Agent 能够为商户策划个性化的营销活动，如优惠券发放、会员活动、节日促销等。并对营销活动的效果进行跟踪和评估，及时调整活动策略。
- **库存与供应链管理**：对于部分实体商户，Agent 还能够协助进行库存管理和供应链优化。通过分析销售数据和库存水平，Agent 能够预测商品的需求趋势，为商户提供库

存补充建议，避免库存积压和缺货情况的发生。

3.3 酒旅业务线大模型 Agent 实践

3.3.1 智能预订助手

智能预订助手为用户提供一站式的酒旅预订服务，涵盖酒店、机票、景点门票等多个方面。大模型 Agent 通过深入理解用户的出行需求，为用户提供个性化的预订方案。

用户可以通过自然语言与智能预订助手交互，告知出行时间、目的地、预算、偏好等信息。Agent 会根据用户的需求，从海量的酒旅资源中筛选出合适的选项，并为用户提供详细的介绍和比较。在用户做出选择后，Agent 会协助完成预订流程，包括填写订单信息、支付等。

同时，智能预订助手还能够为用户提供出行相关的附加服务，如接送机、旅游保险等。在预订过程中，Agent 会及时向用户反馈订单状态和相关信息，确保用户的出行顺利。

3.3.2 行程规划 Agent

行程规划 Agent 能够根据用户的出行需求和偏好，为用户制定详细的行程计划。用户只需告知目的地、出行时间、兴趣爱好等信息，Agent 就会综合考虑景点开放时间、交通状况、餐饮住宿等因素，为用户规划出合理的行程路线。

在行程规划过程中，Agent 会与用户保持交互，根据用户的反馈和需求变化，及时调整行程计划。同时，Agent 还能够为用户提供景点介绍、美食推荐、交通指南等相关信息，帮助用户更好地了解和体验目的地。

行程规划 Agent 还具备一定的灵活性和适应性，能够应对突发情况，如天气变化、景点关闭等，为用户调整行程并提供替代方案。

3.3.3 酒店智能服务 Agent

酒店智能服务 Agent 为入住酒店的用户提供全方位的智能服务，提升用户的入住体验。Agent 通过与酒店的各种系统集成，如客房控制系统、餐饮系统、安保系统等，为用户提供便捷的服务。

用户可以通过语音或文字与酒店智能服务 Agent 交互，提出各种需求，如客房清洁、餐

饮预订、叫醒服务、设施维修等。Agent 会及时响应并安排相关服务。同时，Agent 还能够为用户提供酒店周边的旅游景点、餐饮娱乐等信息推荐。

在用户入住期间，Agent 会实时关注用户的需求和反馈，及时解决用户遇到的问题，确保用户的入住舒适和满意。

3.4 共享单车业务线大模型 Agent 实践

3.4.1 车辆调度与运维 Agent

共享单车业务的核心挑战之一是车辆的合理调度和高效运维。车辆调度与运维 Agent 通过分析用户的出行数据、车辆分布情况、天气等因素，实现车辆的智能调度和运维管理。

Agent 会实时监控各区域的车辆数量和使用情况，当某个区域车辆短缺时，会及时调度附近的车辆前往补充；当某个区域车辆积压时，会安排运维人员将车辆转移到其他需求较大的区域。通过合理的调度，提高了车辆的利用率，减少了用户找车难的问题。

在运维方面，Agent 能够通过车辆传感器收集的数据，实时监测车辆的状态，如电池电量、故障情况等。当发现车辆出现故障或电量不足时，会及时通知运维人员进行维修和充电。同时，Agent 还能够为运维人员规划最优的运维路线，提高运维效率。

3.4.2 用户出行助手

用户出行助手为共享单车用户提供便捷的出行服务。用户可以通过 APP 与出行助手交互，获取附近的车辆信息、推荐最佳骑行路线等。

大模型 Agent 能够根据用户的历史出行数据和当前位置，预测用户的出行目的地，并为用户推荐合适的车辆和骑行路线。在骑行过程中，Agent 还能够为用户提供实时的路况提醒、天气预警等信息，保障用户的骑行安全。

当用户遇到问题时，如车辆故障、还车困难等，用户出行助手会及时响应并提供解决方案。同时，Agent 还能够收集用户的反馈和建议，为共享单车业务的优化和改进提供依据。

第四章 美团大模型 Agent 开发流程

4.1 需求分析与场景定义

需求分析与场景定义是大模型 Agent 开发的首要环节，其质量直接影响后续的开发过程和最终效果。在这一环节，需要深入了解业务需求和用户需求，明确 Agent 的应用场景和目标。

首先，与业务方进行充分沟通，了解业务的痛点和需求，确定 Agent 需要解决的核心问题。通过业务调研、用户访谈、数据分析等方式，收集相关信息，梳理业务流程和规则。然后，根据收集到的信息，定义具体的应用场景。明确 Agent 在场景中的角色、交互对象、任务目标和约束条件。对场景进行详细的描述和分析，包括用户的行为模式、数据特点、业务规则等。

在需求分析和场景定义过程中，需要充分考虑场景的复杂性和多样性，确保 Agent 能够适应不同的情况。同时，要对需求进行优先级排序，确定核心功能和次要功能，为后续的开发规划提供依据。

4.2 数据准备与预处理

数据是大模型 Agent 训练和运行的基础，数据准备与预处理的质量对 Agent 的性能至关重要。在这一环节，需要收集、清洗、标注和转换数据，为模型训练和推理提供高质量的数据支持。

数据收集方面，整合美团内部各业务线的相关数据，包括用户行为数据、业务交易数据、文本数据、图像数据等。同时，根据场景需求，可能还需要收集外部数据，如天气数据、交通数据等。在数据收集过程中，要确保数据的合法性、合规性和隐私性。

数据清洗是去除数据中的噪声、异常值和重复数据，保证数据的准确性和完整性。通过数据清洗，提高数据的质量，减少对模型训练的干扰。

数据标注是为数据添加标签和注释，以便模型能够理解和学习。根据任务类型的不同，数据标注的方式也有所不同，如分类标注、实体标注、关系标注等。对于自然语言处理任务，还需要进行分词、词性标注等预处理操作。

数据转换是将数据转换为模型能够处理的格式，如将文本数据转换为向量表示，将图像数据转换为像素矩阵等。同时，还需要对数据进行归一化、标准化等处理，以提高模型的训练效果。

4.3 模型选型与微调

根据需求分析和场景定义的结果，结合数据特点和业务要求，选择合适的基础模型进行开发。美团的龙猫大模型（LongCat-Flash-Chat）是一个不错的选择，其强大的性能和丰富的功能能够满足大多数业务场景的需求。

在模型选型后，需要对模型进行微调，以使其更好地适应具体的业务场景。微调过程包括以下几个步骤：

- **准备微调数据集：**根据场景需求，从预处理后的数据中筛选出合适的数据集作为微调数据。微调数据集应具有代表性和针对性，能够覆盖场景中的主要任务和情况。
- **确定微调策略：**根据模型类型和数据特点，选择合适的微调策略，如全参数微调、部分参数微调、提示微调等。在微调过程中，需要调整学习率、batch size、训练轮数等超参数，以获得最佳的微调效果。
- **执行微调训练：**使用准备好的微调数据集和确定的微调策略，对模型进行训练。在训练过程中，要实时监控模型的性能指标，如损失函数、准确率等，及时调整训练策略。
- **模型评估与选择：**微调训练完成后，使用验证数据集对模型进行评估。根据评估结果，选择性能最佳的模型作为最终的部署模型。

4.4 Agent 架构设计与实现

Agent 架构设计与实现是将模型与业务逻辑相结合，构建完整的大模型 Agent 系统的过程。在架构设计中，需要考虑 Agent 的模块划分、交互流程、数据流转等方面。

Agent 架构通常包括以下几个核心模块：

- **感知模块：**负责接收和解析外部输入信息，如用户的自然语言请求、环境数据等。感知模块需要具备良好的兼容性和扩展性，能够处理多种类型的输入。
- **决策模块：**基于感知模块获取的信息和模型的推理结果，进行任务规划和决策。决策模块是 Agent 的核心，需要具备强大的逻辑推理和问题解决能力。
- **执行模块：**根据决策模块的指令，调用相关的工具、系统和 API，执行具体的任务。执行模块需要与美团内部的各种系统进行集成，确保任务的顺利执行。

- 记忆模块 :负责存储 Agent 的历史交互数据、任务状态、模型参数等信息。记忆模块能够帮助 Agent 更好地理解用户需求和历史行为 , 提高任务处理的准确性和效率。

在架构实现过程中 , 采用模块化、组件化的设计思想 , 提高系统的可维护性和可扩展性。使用合适的编程语言和框架进行开发 , 如 Python、Java、TensorFlow、PyTorch 等。同时 , 要注重代码的质量和安全性 , 进行严格的代码审查和测试。

4.5 测试与优化

测试与优化是确保大模型 Agent 系统质量和性能的关键环节。在这一环节 , 需要对 Agent 进行全面的测试 , 发现问题并进行优化改进。

测试包括以下几个方面 :

- 功能测试 :验证 Agent 的各项功能是否正常实现 , 是否符合需求规格说明书的要求。

通过设计测试用例 , 模拟各种场景和输入 , 检查 Agent 的输出结果是否正确。

- 性能测试 :评估 Agent 的响应时间、吞吐量、并发处理能力等性能指标。通过压力测试、负载测试等方式 , 检验 Agent 在不同负载情况下的性能表现 , 找出性能瓶颈并进行优化。

- 安全性测试 :检查 Agent 系统是否存在安全漏洞 , 如数据泄露、恶意攻击等。通过安全扫描、渗透测试等手段 , 确保系统的安全性和稳定性。

- 用户体验测试 :邀请真实用户对 Agent 进行使用体验测试 , 收集用户的反馈意见和建议。根据用户反馈 , 优化 Agent 的交互方式、功能设计等 , 提升用户体验。

根据测试结果 , 对 Agent 系统进行优化改进。优化包括模型优化、算法优化、架构优化、代码优化等多个方面。通过不断的测试和优化 , 提高 Agent 的性能和质量 , 确保其能够满足业务需求和用户期望。

第五章 美团大模型 Agent 工程化实践

5.1 开发工具链与平台支持

美团为大模型 Agent 的开发提供了完善的开发工具链和平台支持，以提高开发效率和质量。

在开发工具方面，提供了一系列针对大模型开发的工具，如模型训练框架、数据处理工具、代码调试工具等。模型训练框架支持多种主流的深度学习框架，如 TensorFlow、PyTorch 等，并进行了优化和定制，以适应美团的硬件环境和业务需求。数据处理工具能够高效地处理海量数据，支持数据清洗、转换、标注等操作。代码调试工具提供了强大的调试功能，帮助开发人员快速定位和解决问题。

在开发平台方面，美团构建了统一的大模型开发平台，集成了数据管理、模型训练、模型评估、模型部署等功能。开发人员可以在平台上完成大模型 Agent 的整个开发流程，实现数据共享、模型复用和协同开发。平台还提供了丰富的 API 和 SDK，方便开发人员与其他系统进行集成。

此外，美团还建立了内部的知识库和社区，开发人员可以在其中分享经验、交流问题、获取资源，促进大模型 Agent 技术的发展和應用。

5.2 监控与运维体系

为确保大模型 Agent 系统的稳定运行，美团构建了完善的监控与运维体系。该体系基于 AIOps 理念，通过人工智能技术实现对系统的智能化监控和运维。

监控体系主要包括以下几个方面：

- **指标监控**：实时采集大模型 Agent 系统的各项指标，如 CPU 使用率、内存占用、响应时间、错误率等。通过设置阈值和告警规则，当指标超过阈值时及时发出告警，通知相关人员进行处理。
- **日志监控**：收集和分析系统的日志数据，包括模型训练日志、推理日志、业务日志等。通过日志分析，能够及时发现系统中的异常情况和问题，为故障排查提供依据。
- **链路追踪**：对大模型 Agent 系统的调用链路进行追踪，记录每个请求的处理过程和耗时情况。通过链路追踪，能够快速定位性能瓶颈和故障点，提高系统的可观测性。

运维体系则包括故障发现、告警触达、故障定位和故障恢复等环节。基于 AIOps 的故

障发现系统 Horae 能够自动发现系统中的故障和异常，通过异常检测算法从海量的时序数据中识别出异常模式。告警触达系统对告警事件进行收敛、聚合和抑制，减少告警风暴和误告。故障定位系统通过数据收集、根因分析、关联分析等手段，快速定位故障根因。故障恢复系统则通过流量切换、预案执行、降级等方式，及时恢复系统功能，减少业务损失。

5.3 安全与合规保障

安全与合规是大模型 Agent 开发和应用的重要前提。美团高度重视大模型 Agent 的安全与合规问题，采取了一系列措施确保系统的安全性和合规性。

在数据安全方面，美团建立了严格的数据安全管理制度，对数据的收集、存储、使用、传输等环节进行全程管控。采用数据加密、访问控制、脱敏等技术手段，保护用户数据的隐私和安全。同时，定期进行数据安全审计和风险评估，及时发现和解决数据安全问题。

在模型安全方面，龙猫大模型具有良好的安全性能，对有害信息识别准确率达 83.98%、隐私保护相关任务准确率达 93.98%，尤其在“违法内容识别”上达 91.24%。在模型训练和部署过程中，采取了一系列安全措施，如模型加密、防篡改、权限管理等，防止模型被恶意攻击和滥用。

在合规方面，美团严格遵守国家相关法律法规和行业规范，如《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等。在大模型 Agent 的开发和应用过程中，确保符合数据收集、使用、隐私保护等方面的要求，定期进行合规检查和评估。

第六章 美团大模型 Agent 评估与迭代

6.1 评估指标体系

建立科学合理的评估指标体系是衡量大模型 Agent 性能和效果的关键。美团根据不同的业务场景和任务类型，构建了多维度的评估指标体系。

在功能性指标方面，主要包括任务完成率、准确率、召回率等。任务完成率衡量 Agent 成功完成任务的比例；准确率衡量 Agent 输出结果的正确程度；召回率衡量 Agent 能够覆盖的相关结果的比例。

在性能指标方面，包括响应时间、吞吐量、并发处理能力等。响应时间是指 Agent 从接

收请求到返回结果的时间 ;吞吐量是指单位时间内 Agent 能够处理的请求数量 ;并发处理能力是指 Agent 同时处理多个请求的能力。

在用户体验指标方面 ,包括用户满意度、交互流畅度、易用性等。通过用户调研、问卷调查、用户行为分析等方式收集用户反馈 ,评估用户对 Agent 的满意度和使用体验。

在业务价值指标方面 ,根据不同的业务线设定相应的指标 ,如外卖业务的配送时长缩短率、订单异常处理效率提升率 ;到店业务的用户转化率、商户销售额增长率等。通过这些指标评估大模型 Agent 对业务的实际贡献。

6.2 A/B 测试方法

A/B 测试是评估大模型 Agent 效果的重要方法 ,通过对比不同版本的 Agent 在实际业务场景中的表现 ,选择最优版本进行推广。

美团的 A/B 测试平台采用了基于域和层的流量分配策略。根据实验类型 ,将流量分成可重叠的多个 Layer ,如 UI 层、推荐算法层、广告算法层等。各个层之间的实验结果和观测指标独立 ,不存在相互影响。

在流量分配上 ,以用户维度进行分流 ,保证单一用户每次进来看到的是相同实验版本 ,避免用户体验不一致。通过 Hash 算法 ,将用户唯一标示、策略 ID、Layer ID 等因素组合计算 HashID ,用户根据 HashID 随机均匀地分配到不同的实验组。

在实验过程中 ,实时收集和分析实验数据 ,包括核心指标和辅助指标。通过统计学方法进行显著性检验 ,计算 P-value ,判断不同版本之间的差异是否具有统计显著性。根据实验结果 ,决定是否推广某个版本或进行进一步的优化。

6.3 迭代优化策略

大模型 Agent 的迭代优化是一个持续不断的过程 ,基于评估结果和用户反馈 ,不断改进 Agent 的性能和功能。

迭代优化策略主要包括以下几个方面 :

- **数据驱动优化** :通过分析大量的用户交互数据和业务数据 ,发现 Agent 存在的问题和不足 ,针对性地进行优化。例如 ,根据用户的错误反馈 ,优化模型的理解能力 ;根据业务数据的变化 ,调整 Agent 的决策策略。
- **模型持续训练** :随着新数据的不断积累 ,定期对大模型进行重新训练和微调 ,以适

应新的业务场景和数据分布。同时，引入新的训练方法和技术，不断提升模型的性能。

- **功能迭代升级**：根据业务需求和用户反馈，不断增加新的功能和特性，优化 Agent 的交互方式和用户体验。例如，增加新的工具调用能力，优化语音交互的准确性和流畅度。
- **快速验证与反馈**：采用敏捷开发的方式，快速迭代开发新版本的 Agent，并通过 A/B 测试等方式进行验证。及时收集用户反馈和实验数据，根据反馈结果调整优化方向。通过持续的迭代优化，确保大模型 Agent 能够不断适应业务的发展和用户的需求变化，保持其竞争力和价值。

第七章 美团大模型 Agent 最佳实践与避坑指南

7.1 跨业务线通用原则

在美团各业务线的大模型 Agent 实践中，积累了一些跨业务线通用的最佳实践原则，这些原则能够指导大模型 Agent 的开发和应用。

- **以业务价值为导向**：大模型 Agent 的开发和应用必须紧密围绕业务目标，以解决业务痛点、提升业务效率和用户体验为出发点。在技术选型和功能设计上，要充分考虑业务的实际需求和价值。
- **数据质量是基础**：高质量的数据是大模型 Agent 成功的关键。要重视数据的收集、清洗、标注和预处理工作，确保数据的准确性、完整性和相关性。同时，要建立数据质量管理机制，持续监控和提升数据质量。
- **渐进式落地**：大模型 Agent 的落地应采取渐进式的策略，从简单场景入手，逐步扩展到复杂场景。在每个阶段，要进行充分的测试和验证，确保 Agent 的稳定性和有效性。同时，要根据反馈及时调整和优化。

- 人机协同 :大模型 Agent 并非要完全替代人工 ,而是要实现人机协同 ,提高工作效率和质量。在设计 Agent 时 ,要明确人机分工 ,让 Agent 处理重复性、规律性的任务 ,而人类则专注于复杂问题的解决和决策。
- 持续学习与进化 :大模型 Agent 应具备持续学习和进化的能力 ,通过不断地与业务场景交互和数据积累 ,优化自身的模型和策略。要建立学习机制 ,让 Agent 能够从错误中学习 ,不断提升性能。

7.2 常见问题与解决方案

在大模型 Agent 的开发和应用过程中 ,会遇到各种常见问题 ,以下是一些典型问题及相应的解决方案。

- 模型幻觉问题 :模型幻觉是指大模型生成的结果与事实不符。解决方案包括 :增加高质量的训练数据 ,特别是事实性数据 ;采用知识蒸馏、检索增强生成 (RAG) 等技术 ,引入外部知识来辅助模型生成 ;在模型输出后进行事实性校验 ,过滤掉错误信息。
- 响应延迟问题 :响应延迟会影响用户体验和业务效率。解决方案包括 :优化模型结构 ,采用轻量化模型或模型压缩技术 ;优化推理引擎 ,提高推理速度 ;采用缓存机制 ,缓存常见问题的答案和中间结果 ;合理分配计算资源 ,确保高并发场景下的性能。
- 数据安全与隐私问题 :数据安全和隐私保护是至关重要的。解决方案包括 :采用数据加密、脱敏等技术 ,保护敏感数据 ;建立严格的数据访问控制机制 ,限制数据的使用范围 ;遵守相关法律法规和行业规范 ,确保数据的合法使用 ;定期进行数据安全审计和风险评估。
- 多 Agent 协同冲突问题 :在多 Agent 协同场景中 ,可能会出现目标不一致、资源竞争等冲突问题。解决方案包括 :设计合理的多 Agent 协同策略 ,明确各 Agent 的角色和职责 ;建立 Agent 之间的通信和协调机制 ,实现信息共享和冲突协商 ;采用分布式

决策算法，确保多 Agent 系统的整体优化。

- 用户体验不佳问题：用户体验不佳可能表现为交互不自然、理解不准确等。解决方案包括：优化自然语言处理能力，提高模型的理解和表达能力；改进交互设计，使 Agent 的交互方式更符合用户习惯；加强用户反馈收集和分析，针对性地优化 Agent 的功能和性能。

7.3 成功案例分享

7.3.1 外卖智能调度系统优化案例

外卖智能调度系统是美团外卖业务的核心系统之一，通过引入大模型 Agent 技术，对调度系统进行了优化升级。

在优化前，调度系统在面对复杂路况、突发订单高峰等情况时，订单分配和路径规划的准确性和效率有待提升。通过引入大模型 Agent，利用其强大的数据分析和决策能力，实现了更精准的订单指派和路径规划。

大模型 Agent 能够实时分析骑手的位置、负载、历史表现等数据，结合订单的地址、商家出餐速度、路况等因素，快速做出最优的调度决策。同时，Agent 还能够根据实时数据动态调整调度策略，应对突发情况。

优化后，外卖配送时长平均缩短了 5 分钟，骑手的工作效率提升了 15%，用户满意度提高了 10%。

7.3.2 到店智能推荐系统升级案例

到店智能推荐系统通过大模型 Agent 技术的应用，实现了推荐效果的显著提升。

升级前，推荐系统主要基于传统的协同过滤和规则引擎，推荐结果的个性化程度和准确性不够高。引入大模型 Agent 后，系统能够更深入地理解用户的需求和偏好，结合商户的特征和口碑等信息，提供更精准的推荐。

大模型 Agent 通过分析用户的历史消费数据、搜索记录、评价等信息，构建用户画像和兴趣模型。根据用户的实时行为和场景，动态调整推荐策略，为用户推荐最符合其需求的商户和服务。

升级后，到店业务的用户点击率提升了 20%，转化率提升了 15%，商户的销售额平均增长了 12%。

第八章 未来展望与技术演进

8.1 美团大模型 Agent 技术发展方向

未来，美团大模型 Agent 技术将朝着以下几个方向发展：

- **更强的智能体能力**：不断提升大模型 Agent 的自然语言理解、任务规划、工具调用、环境交互等核心能力，使其能够处理更复杂、更开放的业务场景。通过引入多模态技术，实现文本、图像、语音等多种数据类型的融合处理，提升 Agent 的感知和理解能力。
- **更高的效率和性能**：持续优化模型架构和训练策略，提高模型的推理速度和吞吐量，降低计算成本。研究轻量化模型和模型压缩技术，使大模型 Agent 能够在资源受限的设备上高效运行。
- **更好的安全性和可靠性**：加强模型的安全防护能力，提高对恶意攻击和有害信息的识别和抵御能力。建立更完善的模型评估和验证体系，确保模型的可靠性和稳定性。
- **更广泛的业务覆盖**：将大模型 Agent 技术拓展到更多的业务领域和场景，如生鲜零售、社区服务、金融科技等。实现跨业务线的协同和联动，为用户提供更全面、一体化的服务。
- **更深度的人机协同**：进一步探索人机协同的模式和机制，实现人类与 Agent 的无缝协作。通过 Agent 辅助人类决策、自动化处理繁琐任务，提高工作效率和质量，同时人类为 Agent 提供指导和监督，确保 Agent 的行为符合预期。

8.2 面临的挑战与应对策略

在美团大模型 Agent 技术的发展过程中,还面临着一些挑战,需要采取相应的应对策略。

- **技术瓶颈**:大模型的训练和推理仍然面临着计算资源消耗大、模型解释性差等技术瓶颈。应对策略包括:加大技术研发投入,探索新的模型架构和训练方法;加强与学术界和产业界的合作,引进先进技术和理念;建立技术攻关团队,针对关键技术难题进行突破。

- **数据质量和数量**:随着业务的发展,对数据的质量和数量提出了更高的要求。但数据收集难度加大、数据标注成本高、数据隐私保护严格等问题制约了数据的获取和利用。应对策略包括:优化数据收集和清洗流程,提高数据质量;研究半监督学习、无监督学习等技术,减少对标注数据的依赖;加强数据隐私保护技术的研究和应用,在合规的前提下充分利用数据。

- **人才短缺**:大模型 Agent 技术涉及多个学科领域,需要具备跨学科知识和实践经验的复合型人才。目前,相关人才短缺成为制约技术发展的重要因素。应对策略包括:加强内部人才培养,通过项目实践和培训提升员工的技术能力;引进外部优秀人才,建立多元化的人才团队;与高校和科研机构合作,开展人才培养和科研合作项目。

- **伦理和社会问题**:大模型 Agent 的广泛应用可能引发一些伦理和社会问题,如算法偏见、就业影响等。应对策略包括:建立伦理审查机制,确保技术的发展符合伦理规范;加强对算法的公平性和透明度研究,减少算法偏见;积极探索人机协同的新模式,创造新的就业机会。

8.3 结语

美团大模型 Agent 技术的发展和应用,为美团的业务创新和发展注入了强大的动力。通过不断的技术创新和实践探索,美团在大模型 Agent 领域取得了一系列成果,为用户、商户和社会创造了价值。

未来,美团将继续秉持“零售 + 科技”的战略,加大在大模型 Agent 技术领域的投入,

不断提升技术水平和应用效果。同时，将积极应对技术发展面临的挑战，加强与各方的合作与交流，共同推动大模型 Agent 技术的健康发展。

我们相信，随着大模型 Agent 技术的不断成熟和完善，它将在美团的业务发展中发挥越来越重要的作用，为实现“帮大家吃得更好，生活更好”的使命提供更有力的支撑。

第三章 美团各业务线大模型 Agent 实践（续）

3.5 生鲜零售业务线大模型 Agent 实践

3.5.1 智能选品与库存 Agent

生鲜零售业务具有时效性强、损耗率高的特点，智能选品与库存 Agent 通过融合多模态数据处理技术，实现了生鲜商品全生命周期的智能化管理。该 Agent 每日分析超过 10 万种生鲜商品的历史销售数据、时令节气、区域消费偏好、天气变化等 20 余类因素，生成动态选品方案。

在选品决策中，Agent 能够识别商品的图像特征（如新鲜度、规格）并关联用户评价文本情感分析，精准预测商品受欢迎程度。例如，在台风预警期间，系统会自动上调方便食品、饮用水等应急商品的备货权重，并降低易腐蔬果的采购量。通过这种动态调整，试点区域的生鲜损耗率降低了 18%，库存周转率提升了 25%。

库存管理方面，Agent 与冷链物流系统实时联动，当监测到某类商品库存低于安全阈值时，会自动触发补货流程，同时考虑供应商的实时产能和物流时效，生成最优采购方案。对于临期商品，系统会结合用户价格敏感度模型，自动生成促销建议并推送至 APP 首页。

3.5.2 履约配送优化 Agent

生鲜配送面临“最后一公里”的温度控制、时效保障双重挑战。履约配送优化 Agent 通过融合 IoT 设备数据与大模型决策能力，构建了智能化的配送体系。Agent 为每个骑手配备的智能保温箱安装了温度传感器，实时监测箱内环境并动态调节制冷设备，确保生

鲜商品全程处于最佳保存温度。

在路径规划上，Agent 不仅考虑常规的距离因素，还创新性地引入 "鲜度衰减系数"，对高时效要求的商品（如海鲜、乳制品）优先规划配送路线。通过历史数据训练，系统能够精准预测不同温度下各类商品的品质变化曲线，当预计送达时品质可能下降超过阈值时，会自动触发预警并协调骑手优先配送。

实践数据显示，该 Agent 部署后，生鲜订单的准时送达率提升至 98.7%，商品损坏投诉率下降了 42%，客户复购率提高了 15%。

3.5.3 会员权益智能管理 Agent

生鲜零售业务的会员体系具有高频互动、权益复杂的特点。会员权益智能管理 Agent 通过构建用户分层模型，实现了权益服务的精准化运营。Agent 会分析会员的消费频次、客单价、品类偏好等数据，自动划分会员等级并匹配相应权益。

在会员互动中，Agent 能够理解自然语言请求并提供个性化服务，如 "帮我预约周末的有机蔬菜配送"、"我的会员日优惠券快过期了" 等需求，系统会自动完成预约操作或推荐最优用券方案。对于高价值会员，Agent 会定期生成专属优惠组合，并结合季节特点推送健康饮食建议，提升会员活跃度和忠诚度。

通过该 Agent 的运营，生鲜会员的月均消费频次提升了 30%，权益核销率提高了 22%，会员留存率较传统运营模式增长了 17%。

第二章 美团大模型 Agent 技术架构（续）

2.4 龙猫大模型架构升级（2024）

针对业务场景的不断深化，龙猫大模型在 2024 年进行了重要架构升级，推出 LongCat-Flash-Chat V2 版本。新版本保留了 5600 亿参数的 MoE 架构基础，重点优化了三项核心能力：

动态专家选择机制实现了计算资源的智能调度。通过引入强化学习算法，模型能够根据任务复杂度和实时资源负载，动态调整激活的专家数量（186 亿 - 420 亿参数 / 词元）。在简单的订单查询场景中，系统仅激活基础专家模块；而在复杂的行程规划任务中，则会调用全部专家资源。这种弹性调度使推理效率提升了 40%，在保证服务质量的前提下降低了 35% 的计算成本。

多模态融合模块的加入扩展了模型的感知能力。新增的视觉编码器能够处理商品图片、门店环境、菜品外观等图像数据，与文本理解模块形成跨模态注意力机制。在生鲜零售场景中，模型可以同时分析商品图片的新鲜度特征和用户评价文本，生成更精准的商品推荐。该模块采用对比学习训练策略，在美团内部多模态基准测试中取得 91.3% 的准确率。

边缘 - 云端协同推理架构解决了终端设备的算力限制。通过模型蒸馏技术，生成轻量级终端模型(参数规模降至 28 亿)，负责实时响应本地化请求(如骑手终端的语音指令识别)；复杂任务则上传至云端大模型处理，两者通过联邦学习机制保持知识同步。这种架构使骑手智能助手的离线响应率提升至 85%，平均交互延迟降低至 200ms 以内。

2.5 模型训练效率优化

2024 年训练体系升级采用了混合精度训练与分布式优化相结合的策略，在保持模型质量的前提下，将 LongCat-Flash-Chat V2 的训练周期缩短至 22 天(较上一代减少 27%)，单万亿词元训练成本降低至 0.32 美元。

创新的梯度压缩技术将通信量减少了 60%，通过低秩分解和量化处理，在不损失训练精度的前提下，显著提升了分布式训练效率。自适应学习率调度器能够根据不同训练阶段自动调整学习率，预训练阶段采用较大学习率快速收敛，微调阶段则自动降低学习率以优化参数。

针对生鲜零售等垂直领域的训练需求，开发了领域知识注入机制。通过构建生鲜商品知识图谱(包含 12 万实体和 35 万关系)，采用知识增强预训练方法，使模型在生鲜术语理解、食品安全规范掌握等方面的准确率提升了 28%。

第五章 美团大模型 Agent 工程化实践(续)

5.4 绿色 AI 工程实践

美团将可持续发展理念融入大模型 Agent 的工程化实践，构建了环境友好型的 AI 技术体系。通过优化模型架构和训练策略，LongCat-Flash-Chat V2 的单任务能耗较上一代降低了 27%，相当于每年减少碳排放约 4200 吨。

在推理环节引入了动态算力调度系统，该系统会根据业务峰谷规律和可再生能源供电情况，智能调整计算资源分配。例如，在夜间风电、光伏供电充足时段，系统会集中处理模型更新和离线计算任务；而在用电高峰时段，则通过模型压缩和缓存机制减少算力消

耗。这种调度策略使数据中心的绿电使用率提升至 38%。

骑手配送路径优化 Agent 新增了碳排放因子考量,在规划路线时不仅考虑距离和时间,还会计算不同配送方案的碳排放量,优先推荐环保路线。系统通过融合交通流量数据和新能源汽车充电网络信息,为骑手规划最优配送路线,试点区域的单车百公里碳排放降低了 15%。

建立了 AI 系统碳足迹追踪平台,能够实时监测大模型训练、推理全流程的能源消耗和碳排放数据,并生成可视化报告。该平台已纳入美团 ESG 评估体系,成为技术创新与绿色发展协同推进的重要支撑。

第六章 美团大模型 Agent 评估与迭代 (续)

6.4 多维度评估体系升级

2024 年评估体系新增了环境友好度和社会价值两个维度,形成更全面的评估框架。环境友好度指标包括模型训练能耗、推理阶段碳排放量、资源利用率等量化数据,通过开发的 EcoScore 评分模型进行综合评价,确保技术与绿色低碳目标一致。

社会价值评估聚焦于大模型 Agent 对就业质量的提升作用,包括骑手收入增长幅度、商户运营效率提升带来的就业稳定性、老年人等特殊群体的服务可及性等指标。在老年用户服务场景中,通过语音交互优化和简化操作流程,60 岁以上用户的订单完成率提升了 32%。

针对多 Agent 协同场景,开发了协同效率评估基准 CoopBench,从任务分配合理性、资源冲突解决能力、全局优化效果等维度进行量化评估。在外卖与共享单车 Agent 的协同调度测试中,该基准成功识别出 3 类协同瓶颈,通过算法优化使极端天气下的运力利用率提升了 20%。

6.5 智能迭代优化系统

基于强化学习的智能迭代系统实现了模型优化的自动化闭环。该系统会实时收集各业务线 Agent 的运行数据,通过多目标优化算法确定迭代方向,并自动生成微调方案。对于高频出现的问题(如特定场景的语义理解错误),系统能够快速定位原因并推荐解决方案,将模型迭代周期从平均 14 天缩短至 7 天。

开发了用户反馈深度挖掘模块,通过情感分析和意图识别技术,从海量用户评价中提取有价值的改进建议。例如,系统发现用户对“生鲜配送时间预估不准”的抱怨集中在雨

天场景后，自动触发了该场景下的模型专项优化，使相关投诉下降了 53%。

建立了模型健康度监测体系，通过设定关键指标阈值（如准确率、响应时间、资源消耗），实时监控模型性能变化。当检测到性能异常时，系统会自动启动诊断流程，判断是数据漂移、概念偏移还是系统故障导致，并采取相应的恢复措施，保障模型服务的稳定性。

第七章 美团大模型 Agent 最佳实践与避坑指南（续）

7.4 多 Agent 协同优化案例

7.4.1 暴雨天气运力协同案例

在极端天气条件下，外卖配送与共享单车业务常面临运力资源竞争问题。通过部署多 Agent 协同决策系统，实现了跨业务线的资源优化配置。该系统包含三个核心机制：动态优先级协商机制能够根据实时需求调整运力分配。暴雨来临时，系统会自动提高外卖 Agent 的优先级系数（从 1.0 升至 1.5），确保生活物资配送；同时降低共享单车 Agent 的调度强度，减少骑手暴露风险。当外卖订单压力缓解后，系统会逐步恢复平衡状态。

资源共享池机制将闲置运力转化为应急储备。当检测到某区域外卖运力紧张时，系统会协调附近的共享单车运维车辆参与应急配送，通过临时权限开放和标准化接口，实现跨业务系统的无缝协同。在 2024 年夏季暴雨期间，该机制使订单延误率降低了 37%。

风险共担算法通过利益均衡机制保障多方权益。系统会根据协同贡献度动态调整各业务线的成本分摊和收益分配，确保协同决策的长期可行性。实践表明，该算法有效解决了协同中的“搭便车”问题，使跨业务协同的参与度保持在 90% 以上。

7.4.2 生鲜与外卖库存协同案例

生鲜零售与外卖业务的库存共享 Agent 实现了商品资源的高效利用。当某区域外卖商家的某类商品售罄时，Agent 会自动查询附近生鲜仓的库存情况，在满足生鲜订单需求的前提下，协调调拨部分商品补充外卖库存，实现了库存共享和损耗降低的双重目标。该系统采用分布式锁机制确保库存数据一致性，通过预测算法提前识别潜在的库存缺口，实现主动协同。在试点区域，商品缺货率下降了 28%，而生鲜仓的临期商品利用率提升

了 40%，形成了双赢局面。

7.5 前沿技术落地陷阱与规避

7.5.1 多模态融合陷阱

在生鲜商品识别等场景中，盲目追求多模态融合可能导致系统复杂度飙升而收益有限。某试点项目曾因强行融合图像、文本、传感器数据，导致模型推理延迟增加 3 倍，而准确率仅提升 2%。解决方案是：

建立模态必要性评估矩阵，从数据质量、贡献度、成本三个维度评估是否需要引入该模态；采用渐进式融合策略，先实现单模态最优，再逐步引入其他模态；设置融合增益阈值，当新增模态的边际效益低于 10% 时停止融合。

7.5.2 过度自动化风险

在商户运营助手的开发中，曾出现过度追求自动化导致的“算法暴政”问题，系统自动生成的营销方案因缺乏人工干预而出现不符合实际场景的情况。规避策略包括：

设置关键决策人工审核节点，涉及大额优惠、重要规则变更等必须经过人工确认；建立算法可解释性报告机制，让商户理解系统决策依据；保留人工否决权和调整通道，允许商户根据实际情况修改系统方案。

第八章 未来展望与技术演进（续）

8.4 2025-2027 技术发展路线图

美团大模型 Agent 技术将沿着“通用智能 + 垂直深化”的双轨路线发展，重点推进五大技术方向：

多模态交互升级计划在 2025 年实现 AR 导航与语音交互的深度融合，骑手通过 AR 眼镜即可获取可视化配送路线，系统能够理解复杂环境描述（如“在超市第三个货架取货”）并生成精准指引。该技术将使配送效率再提升 15%。

联邦学习规模化应用将在 2026 年覆盖 80% 的商户数据处理场景，通过隐私保护计算

技术，在不获取原始数据的情况下实现模型优化，解决中小商户的数据安全顾虑。预计该技术将使商户参与智能化改造的意愿提升 40%。

通用智能体架构计划在 2027 年实现跨业务线能力复用，开发具备迁移学习能力的基础 Agent，能够快速适应新业务场景（如社区团购、家政服务）。该架构将使新业务智能化部署周期从 3 个月缩短至 1 个月。

绿色 AI 技术将持续优化模型效率，目标是到 2027 年将大模型训练的单位能耗再降低 50%，推理阶段全面实现 100% 绿电支持，打造行业领先的低碳 AI 技术体系。

人机协同操作系统将构建人类与 Agent 的新型协作模式，通过意图预测和能力互补，使人工干预效率提升 50%，实现“人类指导 + Agent 执行”的最优分工。

8.5 生态共建与社会责任

美团将开放大模型 Agent 技术能力，构建“平台 + 商户 + 开发者”的协同生态。计划在 2025 年发布 Agent 开发平台，提供标准化接口和工具链，支持第三方开发者为特定场景开发定制化 Agent 应用，如特色餐饮推荐、地方特产营销等。

针对老年人、残障人士等特殊群体，将开发适老化 Agent 服务，优化语音交互、简化操作流程，解决数字鸿沟问题。目标是到 2026 年实现老年用户智能服务使用率达 60% 以上。

在食品安全领域，将深化大模型 Agent 的应用，通过图像识别、文本分析等技术，实现从食材采购到餐桌的全链条安全监控，为消费者提供更可靠的食品安全保障。

通过技术创新与生态共建，美团大模型 Agent 将不断拓展应用边界，在创造商业价值的同时，为社会可持续发展贡献科技力量，真正实现“帮大家吃得更好，生活更好”的企业使命。