由大语言模型驱动的餐次级个性化营养管理闭环多智能体系统

Muqing Xu*工程数学系
布里斯托大学 英国布里斯托
hu24149@bristol.ac.uk

arXiv:2601.04491v1 [cs.AI] 8 Jan 2026

摘要—**个性化营养管理旨在根据个人的摄入量和表型提供量身定制的膳食指导，但现有大多数系统将饮食记录、营养素分析和推荐分开处理。我们介绍了一款新一代移动营养助手，它将基于图像的餐食记录与由大语言模型驱动的多智能体控制器相结合，提供餐次级闭环支持。该系统协调视觉、对话和状态管理智能体，根据照片估算营养素，并更新每日摄入预算。随后，它会根据用户偏好和饮食限制调整下一餐的膳食计划。使用 SNAPMe 餐食图像和模拟用户进行的实验表明，该系统在营养素估计、个性化菜单生成和高效任务规划方面表现出竞争力。这些发现证明了多智能体大语言模型控制用于个性化营养的可行性，并揭示了图像微量营养素估计及大规模真实世界研究方面仍面临的挑战。**

索引词—**个性化营养管理，多智能体系统，多模态大语言模型，人工智能智能体，闭环流程**

I. 引言

个性化营养在医疗保健和疾病预防领域日益受到关注。它能够改善健康状况，并帮助预防和控制糖尿病、肥胖等慢性病。然而，传统的膳食评估方法往往存在局限。这些方法通常依赖自我报告或营养师访谈 [1], , 例如24小时膳食回顾（24HR）、饮食日志（DR）和食物频率问卷（FFQ）。对于大多数未接受营养学培训的人来说，这些方法中报告的食物种类和份量很大程度上取决于用户的主观判断，因此可能导致结果出现偏差和不准确。移动健康（mHealth）有望为此提供解决方案。

移动健康（mHealth）通常指由移动设备、无线传感器和通信技术支持的医疗保健与公共卫生实践 [2]。在营养管理中，移动健康极大地提升了膳食评估和干预的效果。与传统方法相比，移动应用可以通过食物图像识别、条形码扫描以及连接营养数据库，实现更快速、准确的监测 [3] [4]。此外，借助人工智能和机器学习，它们还可以根据个人特征和健康需求提供个性化营养建议，为慢性病预防和健康维护开辟了新途径 [5]。

近年来，研究人员开始探索将多智能体系统（MAS）用于营养管理 [6]。在这些系统中，不同的智能体协同完成数据收集、食物识别、方案生成和用户交互等任务 [7]。这种方法具有可扩展、高效和稳健等优势，同时还能为用户提供个性化服务，为下一代个性化营养系统奠定基础。

本文提出了一种基于闭环多智能体设计的餐食级个性化营养指导系统，并配套开发了一款实用的移动应用。该系统根据用户的多模态输入提供个性化饮食建议，并支持动态调整营养计划。我们采用美国农业部（USDA）的膳食参考摄入量（DRIs） [8] [9] [10] 作为营养素摄入量数据库。餐食图像及其营养数据来自餐食照片营养素评估调查（SNAPMe） [11] 数据库，测试则在 Android Studio模拟器上进行。第II节回顾与本文目标相关的工作。第III节介绍用于多模态输入处理的系统架构与设计，包括模型选择、智能体之间的通信以及移动应用开发。第IV节描述实验设置，并展示用于评估系统性能的数据分析。最后，第V节总结全文。

II. 相关工作

近年来，人工智能（AI）迅速发展，为解决营养管理中长期存在的难题提供了前景广阔的方案。其中，基于视觉的膳食评估已成为一种广泛应用的方法。这一进展很大程度上得益于大规模膳食数据集的创建与整理，它们是人工智能驱动的评估系统的重要训练资源。知名实例包括 Food2K 数据集 [12], Recipe1M+ 数据集 [13], 和 Hd-epic 数据集 [14]。基于丰富的数据集，研究人员开发了许多用于基于视觉的膳食评估的人工智能算法和工具

例如，自动食物识别 [15]和食材识别 [16],食物分割 [17], 食物体积估计 [18], 食谱检索 [19]或生成 [20],, 以及利用机器学习智能传感器检测自由生活环境中的进食事件 [21]和行为 [22] ，从而为每个人提供准确、便捷的日常膳食评估。这些技术极大地改变了营养管理中膳食摄入量的评估方式，使营养生态系统的许多环节更加高效、自动化，也影响了人们管理日常摄入量的方式 [23]。

大语言模型（LLMs）在营养管理中也发挥着日益重要的作用。除作为通用营养助手外，一篇题为《使用多模态ChatGPT进行膳食评估：系统分析》 [24]的论文还研究了多模态大语言模型（MLLMs），尤其是 GPT-4V，在膳食评估方面的潜力。该研究利用可穿戴摄像头收集参与者的摄入数据，并进行食物识别、份量估计和营养分析，随后综合这些结果，以识别营养缺口并推荐食物。研究结果表明，GPT-4V 在这些任务中表现良好，证明了使用多模态大语言模型进行营养管理的可行性。

然而，许多用于膳食评估的模型和算法仍有局限。它们通常只能处理单一、离散的任务，例如识别食物或检索特定菜肴的食谱。这些方法缺乏对膳食知识、日常实践、评估流程，以及个人饮食行为和文化背景差异的全面理解。因此，它们无法实现准确量化个人食物和营养素摄入量这一总体目标，也无法为健康饮食和健康管理提供真正个性化的指导。

尽管近年来出现了多智能体推荐系统 [6] [25],，但它们并未针对个性化营养管理提供专门的流程和动态调整策略。为弥补这一空白，本文提出了一种多智能体工作流：“餐食营养分析—动态营养计划调整—个性化膳食推荐”。

III. 闭环系统架构

A. 整体系统工作流

系统的整体工作流如图 1 所示。每天开始时，系统都会生成一份“每日营养摄入计划”。系统根据用户档案和膳食参考摄入量（DRIs）数据库生成该计划，制定一组每日目标，并提供“剩余限额”状态。当用户发送餐食照片并附上简短描述时，“餐食分析模块”会将其转换为结构化营养数据，并向“用户餐食数据库”写入一条带时间戳记录。每次写入后，系统都会更新计划状态，反映当天已摄入的内容和剩余量。随后，“膳食推荐模块”读取当前计划和最新的摄入汇总，并为下一餐提供推荐。如果没有有效图像或置信度较低，系统会采用纯文本估计，或向用户请求澄清。该循环会在一天中的每餐持续进行，并在次日开始时重置。通过将分析、状态管理和推荐分离，系统确保数据契约清晰，并支持审计；闭环还限制了错误传播，使每项推荐都与最新的摄入量保持一致。

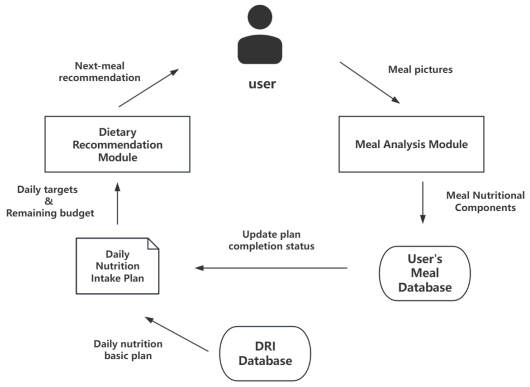


图 1. 系统工作流程概览。

B. 闭环动态调整策略

为更好地支持个性化和动态适应，推荐模块在摄入计划中结合了逐餐调整与逐日调整。在我们提出的膳食推荐和计划更新策略中，下一餐的指导依据当前计划状态，而非固定的每日菜单。系统利用图像和文本中的餐食证据，准确记录实际摄入量，并为下一餐提供可执行建议。每餐结束后，系统汇总实际摄入的食物，并更新当天剩余的营养“预算”。生成推荐前，服务会读取最新的完成状态，并将剩余需求分配到当天剩余的餐次中。随后，系统结合长期记录和用户习惯记录中的偏好与限制条件，推荐最适合填补主要营养素缺口的食物及份量。该设计逐步缩小“每日剩余目标”与“摄入量”之间的差距，从而实现更精准、更个性化的营养管理。

当数据稀疏时，推荐服务仍会提供保守建议。这一设计将每顿饭都视为迈向目标的一小步。与早晨制定一份庞大且固定的计划相比，它减轻了用户负担，提高了依从性。

总之，我们将该任务建模为一个闭环决策过程：每餐都会更新系统状态，并影响下一餐的推荐。此外，每天生成新计划时，计划生成服务会参考前一天的完成情况，并据此调整新计划。餐级和日级闭环调整流程如图2所示。

表 I 智能体的主要职责与大语言模型选择

智能体名称	主要职责	大语言模型选择	选择理由
控制器智能体	规划器和协调器；对请求进行分类 并选择最短且最安全的工作流。	GPT-4o-mini	出色的推理和多步规划能力 同时平衡延迟和成本。
视觉智能体	餐食图片和营养分析器；将 餐食图片转换为结构化营养数据。	GPT-4o	在食物识别和 质量/体积估算方面表现出色；对遮挡具有鲁棒性。
文件智能体	系统状态的唯一编写者和管理员； 更新包含对话内容的文件。	Claude Sonnet 4	更强的指令遵循能力、更好的格式 合规性、更稳定的工具调用、更少的幻觉。
对话智能体	营养顾问；根据计划生成下一餐 基于该计划的推荐。	GPT-4.1-mini	速度快、成本效益高；足以读取小型 数据文件并进行简单的分配/分析。

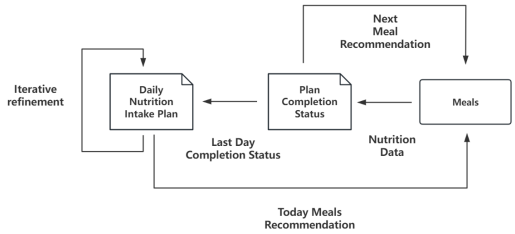


图2. 闭环动态调整策略

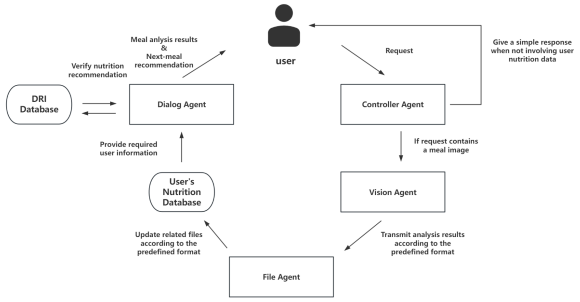


图3. 多智能体系统多模态主工作流概览

C. 多智能体系统架构

我们将主要工作流分解为若干子任务——膳食图像识别、营养分析，以及营养计划的制定与调整——并采用多智能体系统（MAS），由不同的智能体负责一项或多项任务。本文提出的多智能体系统由四个职责明确的智能体组成：控制器智能体、对话智能体、视觉智能体和文件智能体。这四个智能体的职责及模型选择见表 I。视觉智能体的模型选择实验见第IV节。

多智能体系统通过控制器智能体提供了一个面向用户交互的统一入口。每个用户请求都会被封装成一条消息，其中包括用户ID、日期、用餐时间、可选的图像和文本，以及膳食ID。从一开始，我们就要求系统同时回答特殊营养问题和一般问题。为避免浪费令牌，控制器会在摄入时对每条文本进行分类。如果查询属于一般问题，或是不涉及个人数据的简单营养问题，控制器便直接作答。如果问题涉及营养管理，控制器会将任务分解，选择最短且安全的工作流，并安排各个步骤。随后，它会将每个步骤委派给其他智能体。智能体之间使用预定义的载荷格式进行通信，从而确保接口稳定且易于审查。图3展示了多智能体系统多模态主工作流概览。

IV. 实验与结果分析

A. 数据集

膳食参考摄入量（DRIs）是由美国国家科学院食品与营养委员会制定的、以科学为依据的参考值，旨在支持对健康人群进行膳食规划和评估。这些表格由美国农业部国家农业图书馆（NAL）汇编并发布，供从业人员使用。DRIs 包含详细的每日摄入量表，涵盖 40 种必需营养素和元素，分为三类：必需矿物质、维生素和常量营养素。DRI 框架针对不同用途提供了若干数值：平均需要量（EAR）用于评估某个人群的摄入量是否充足；推荐膳食供给量（RDA）用于设定个人目标。对于能量，DRIs 提供估计能量需要量和可接受的常量营养素分布范围。

本文从美国农业部数据库中选取了必需矿物质、维生素和常量营养素的 RDA 表。这些表被合并为一个数据文件，作为规划每日营养素摄入量的依据。RDA 值按生命阶段和性别分层，以反映不同生理需求。在合并之前，我们按照统一规则清理了各表中的数值字段，并调整了部分营养素的单位。经过这一处理，得到三个主键和命名规则相同的干净表。接下来，我们根据类别和生命阶段维度，对这三个表进行了全外连接。合并后，每个生命阶段对应一行，其中并列列出所有营养素。在系统工作流程中，该文件可作为生成每日摄入量目标时可靠的外部参考。

SNAPMe数据库是一个用于基于照片的膳食评估的基准数据集。本研究在自由生活环境中收集餐食照片，并将其与同期的24小时食物记录关联，从而可以将模型输出与结构化自我报告进行比较。该数据集包含3,311份独特的食物

照片，来自95名参与者，涵盖275个进食日。在我们的实验中，我们使用SNAPMe比较不同大语言模型在餐食图像分析方面的表现，并主要通过食物识别、体积和质量估计以及营养素分析的准确性评估其能力。

B. 餐食图像分析能力

鉴于餐食图像分析的准确性是个性化营养管理的关键组成部分，在构建视觉智能体之前，我们对几种主流多模态大语言模型进行了对比评估。我们在 SNAPMe 数据集和带标签的本地餐食照片样本上测试了它们的表现，重点考察其食物类别识别和体积/质量估计能力，以确定最终模型。食物识别测试使用了涵盖 50 种食物的 21 张图像；体积/质量估算测试则使用了 30 张包含参照物的图像和 30 张不包含参照物的图像。结果见表 II 和表 III。

表 II 不同大语言模型的食物识别能力对比

模型	Food 识别准确率	Max 遮挡率[a]	Food 组成准确率
GPT-4o	88%	$>80\%$	$\sim$75%
Qwen QVQ-MAX	72%	60%	$\sim$80%
Gemini 2.5 Flash	92%	$>80\%$	$\sim$80%
Claude Sonnet 4	64%	60%	$\sim$70%

[a] 这表示食物遮挡程度最高。

在食物识别方面，GPT-4o 和 Gemini 2.5 Flash 的整体表现明显优于 Qwen QVQ-MAX 和 Claude Sonnet 4。两者都保持了较高的准确率，并且在严重遮挡（超过 80%）的情况下仍然可靠。Qwen QVQ-MAX 在东亚菜肴上略有优势，但对被遮挡食物的鲁棒性较弱，召回率最高约为 60%。Claude Sonnet 4 在检测时较为保守，通常只报告菜肴的主要食材。这种行为会降低其对组合菜肴或复杂菜肴的召回率。因此，后续体积/质量估算测试仅测试了 GPT-4o 和 Gemini 2.5 Flash。

表 III 食物相关估算的误差范围和平均值

类别	GPT-4o		Gemini 2.5 Flash	
	范围	Mean	范围	Mean
体积估算 [a]	3.5%–12%	7.6%	4.2%–11.7%	7.3%
质量估计 [a]	11.2%–24.5%	17.4%	13.3%–21.9%	16.7%
总能量估算 [a]	12.6%–33.6%	18.7%	13.6%–30.8%	18.2%
体积估算 [b]	4.3%–27.5%	10.7%	6.1%–17.5%	13.8%
质量估计 [b]	15.2%–42.4%	26.7%	14.7%–39.9%	24.1%
总能量估计 [b]	16.4%–49.7%	28.9%	18.5%–45.8%	27.4%

[a] 有参照物；[b] 无参照物。

根据测试结果，Gemini 2.5 Flash 在估计食物质量、体积和能量方面略优于 GPT-4o。然而，由于 Gemini 在测试期间偶尔会拒绝进行体积和质量估计，我们最终选择 GPT 作为驱动视觉智能体的模型，以确保工作流稳定。两种模型在图中有参照物的测试中，表现也明显优于没有参照物的测试。因此，视觉智能体的系统提示词特别强调了使用参照物的重要性：它指示智能体在分析前优先关注用户提示词中提到的任何参照物，或者尝试在图像中寻找常见的参照物。

C. 营养分析能力

我们考虑餐食 $i = 1, \ldots, N$ 和营养素 $j \in \mathcal{J}$。令 $y_{i,j}$ 表示餐食 i 中营养素 j 的真实值含量，$\hat{y}_{i,j}$ 表示预测含量。

定义

$$N_j = \sum_{i=1}^{N} m_{i,j}. \tag{1}$$

对于每种营养素 j，平均绝对误差是在非缺失预测值上计算的：

$$\mathrm{MAE}_j = \frac{1}{N_j} \sum_{i=1}^{N} m_{i,j} \left| \hat{y}_{i,j} - y_{i,j} \right|. \tag{2}$$

覆盖率表示具有非缺失预测值的餐食所占比例：

$$\mathrm{Coverage}_j = \frac{1}{N} \sum_{i=1}^{N} m_{i,j}. \tag{3}$$

在本次测试中，我们从 SNAPMe 数据集中选取了 40 张未包装食品图像，以评估视觉智能体进行营养素分析的能力。营养素分为三组：

- 完整的 40 项。此集合包含营养计划所跟踪的所有营养素。
- 子集1：核心营养素。此子集侧重于直接影响膳食规划和反馈的变量：总能量、蛋白质、碳水化合物、脂肪、总膳食纤维和钠。
- 子集2：微量营养素。此子集涵盖核心集合以外的维生素和矿物质。

结果如表 IV 所示。从测试结果来看，该智能体在营养素分析过程中表现出审慎的保守倾向，而微量营养素往往会增加营养素含量估算的误差。

表 IV 营养分析能力测试结果

字段集	平均绝对误差（千卡）	平均绝对误差（克）	平均绝对误差（毫克）	平均绝对误差（微克）	覆盖率
完整40个字段	65.2	7.4	332	208	0.76
子集1	58.9	6.8	225	—[a]	0.96
子集2	—[a]	—[a]	318	189	0.61

[a] “—”表示不适用。

D. 个性化能力

在个性化实验中，我们通过编辑记录文件更改用户档案（饮食习惯和食物频率），并测试对话智能体是否会调整其推荐。在每日营养计划保持不变的情况下，我们准备了两类记录文件：一类在食物频率上偏向中式菜肴，另一类偏向英式菜肴。图 4 和图 5 展示了推荐结果的对比。输出结果表明，在满足营养目标的同时，菜肴选择会随偏好数据而变化：一类采用中式主食，另一类采用英式主食。结果显示，对话智能体能够利用记录文件中的用户档案实现有效的个性化，同时不影响目标一致性；最终菜单的营养构成保持一致，风格各异，并兼顾用户偏好。在下一项测试中，我们在记录文件的过敏字段中添加了“海鲜过敏”，并使用偏向英式菜肴的记录文件。对话智能体给出的这项推荐如图 6 所示。

For lunch target approximately 25 to 30 g protein and approximately 8 to 12 g fiber within approximately 700 to 900 kcal and at most 800 mg sodium; Next-meal plan: Roast chicken (160 to 180 g) with new potatoes (200 g) and peas (120 g) approximately 30 to 34 g protein, 8 to 10 g fiber, 780 to 860 kcal; Jacket potato (medium) plus baked beans (150 g) plus two boiled eggs plus side salad approximately 28 to 32 g protein, 10 to 12 g fiber, 760 to 880 kcal; Baked cod (180 g) plus wholemeal toast (2 slices) plus vegetable soup (cup) approximately 27 to 31 g protein, 8 to 10 g fiber, 720 to 820 kcal; Alternatives: cod and peas (grilled or baked, small portion) plus wholemeal roll, light or no tartar sauce; Tips: log after lunch for recalculation and drink at least 400 ml water by mid-afternoon.

图 4. 针对英国菜肴偏好的个性化下一餐建议

For lunch target approximately 25 to 30 g protein and approximately 8 to 12 g fiber within approximately 700 to 900 kcal and at most 800 mg sodium; Next-meal plan: Mapo tofu (200 g, reduced oil) plus brown rice (120 g cooked) plus blanched bok choy (150 g) approximately 28 to 32 g protein, 8 to 10 g fiber, 750 to 820 kcal; Steamed sea bass (180 g) plus mixed-grain rice (120 g cooked) plus smashed cucumber (200 g, light soy) approximately 30 to 34 g protein, 7 to 9 g fiber, 720 to 800 kcal; Chicken-broccoli stir-fry (200 g chicken, 200 g veg, low oil) plus buckwheat noodles (90 g cooked) approximately 30 to 35 g protein, 8 to 11 g fiber, 760 to 880 kcal; Alternatives: clear-broth beef noodles (small) plus extra greens plus poached egg; low oil and no chili oil; Tips: log after lunch so I can update dinner, drink at least 400 ml water by mid-afternoon.

图 5. 针对中式菜肴偏好的个性化下一餐建议

For lunch target approximately 25 to 30 g protein and approximately 8 to 12 g fiber within approximately 700 to 900 kcal and at most 800 mg sodium, with a seafood allergy constraint (avoid fish and shellfish and anchovy-based sauces); Next-meal plan: Roast chicken (160 to 180 g) with new potatoes (200 g) and peas (120 g) approximately 30 to 34 g protein, 8 to 10 g fiber, 780 to 860 kcal; Jacket potato (medium) plus baked beans (150 g) plus two boiled eggs plus side salad approximately 28 to 32 g protein, 10 to 12 g fiber, 760 to 880 kcal; Roast beef wholemeal sandwich (2 slices bread, 100 g roast beef) plus low-sodium vegetable soup (cup) plus small jacket potato (150 g) approximately 27 to 31 g protein, 8 to 10 g fiber, 720 to 820 kcal; Alternatives: grilled chicken wrap on wholemeal with lettuce and tomato, or cottage pie (small, lean beef) with peas and carrots; Tips: request no fish stock, no Worcestershire or anchovy sauces, and separate fryers to avoid cross-contact, then log after lunch so I can update dinner.

图 6. 为海鲜过敏用户提供的个性化下一餐建议。

E. 多智能体系统任务规划与完成能力

在本测试中，我们使用规划最优性（路径最优性）来评估任务规划与执行能力。该指标根据智能体针对每项任务的工具调用可执行轨迹计算得出。

表 V 多智能体系统总体测试结果

指标	平均值	最大值	最小值
L_{E2E}	65.4 秒	87.8 秒	53.1 秒
PO	0.75	0.86	0.60

实例。对于任务实例 t，令 $s(t)$ 表示已执行的步骤数，$s^*(t)$ 表示策略定义的最小步骤数。

$$\mathrm{PO}(t) = \frac{s^*(t)}{s(t)}, \qquad \mathrm{PO}(t) \in (0, 1]. \tag{4}$$

较高的路径最优性表示规划方案更接近最短有效工作流。我们根据 Flask API 记录的服务器端时间戳测量延迟。对于任务实例 t，设 $\tau_{\text{in}}(t)$ 为收到请求的时间，$\tau_{\text{out}}(t)$ 为发送最终答案的时间。

$$L_{\mathrm{E2E}}(t) \;=\; \tau_{\text{out}}(t) - \tau_{\text{in}}(t)\;. \tag{5}$$

在这项测试中，我们使用了 50 个复杂度各异的多模态请求，并为每个请求标注了估计的最少步骤数。详细结果见表 V。测试表明，MAS 达成了其功能目标：在约 65 秒内返回结构化、可执行的答案，同时尽可能接近最短有效工作流。与标注的最少步骤数 s^* 相比，这意味着控制器通常每四个最优步骤会额外增加约一步；最坏情况下，路径约为最短路径的 1.67 倍。作为比较，我们以一项相关研究作为基准。首先，导航领域文献建议使用路径长度加权成功率（SPL）作为与路径最优性等效的效率指标。根据 Anderson 等人 [26]，在未知且复杂的环境中达到 0.5 被视为合理目标。按照这一标准，我们的路径最优性更高（0.75）。

我们量化评估：根据近期未满足的残差，次日目标是否朝预期方向调整。

设 $\mathcal{N}$为跟踪的营养素，$\mathcal{T} = \{1, \ldots, T\}$ 为天数。对于$n \in \mathcal{N}$ 和 $t \in \mathcal{T}$，令 $T_n(t)$和 $A_n(t)$分别表示计划目标和实际摄入量。定义

$$E_n^{(k)}(t) = \frac{1}{k} \sum_{i=0}^{k-1} \big(T_n(t-i) - A_n(t-i)\big). \tag{6}$$

$$\Delta T_n(t{+}1) = T_n(t{+}1) - T_n(t). \tag{7}$$

并令 $s \in \{+1, -1\}$ 表示策略方向。我们采用方向一致性（DA）：

$$\Omega = \{(n, t) \mid n \in \mathcal{N}, t \in \{1, \ldots, T-1\}\}, \tag{8}$$

$$\mathrm{DA} = \frac{1}{|\Omega|} \sum_{(n,t)\in\Omega} \mathbf{1}\Big[\mathrm{sign}(\Delta T_n(t{+}1)) = \mathrm{sign}\big(sE_n^{(k)}(t)\big)\Big] \tag{9}$$

我们仅统计满足 $E_n^{(k)}(t) \neq 0$ 的配对。我们在 20 个合成完成情景上进行评估（不足：低于 -50%，接近目标：介于 -5% 和 5% 之间，过量：高于 50%，

表VI 个性化每日计划更新测试结果

方法	DA	95%置信区间
MAS	0.82	[0.78，0.86]
随机基线 0.50	[0.43, 0.57]	
静态方法	0.00	[0.00，0.00]

逐营养素扰动约为 10%），并报告方向准确率的平均值和 95%置信区间。

表VI显示，MAS的表现显著优于两个基线，表明其具备可靠的方向性适应能力；DA仅评估方向。

V. 结论

本文提出了一种大语言模型驱动的多智能体系统架构，并结合闭环动态更新策略，以克服传统营养管理中的若干障碍。实验结果表明，所提出的系统和策略能够有效提供个性化营养指导。我们的研究考察了系统的任务规划与执行、餐食图像分析以及个性化能力。具体而言，我们评估了智能体自主规划工作流的能力，使用规划最优性衡量规划效率，并考察营养分析的完整性和准确性。然而，受时间和成本限制，评估规模相对较小；此外，仅凭图像估算微量营养素仍极具挑战。未来的工作将着重扩大测试范围并提高测试的真实性，开展受控用户研究，以量化其对用户负担、依从性和满意度的影响。我们还将结合图像推理与轻量级检索，在不过度夸大确定性的前提下提升微量营养素覆盖率，使系统更接近专业营养师的表现。近期研究表明，利用大语言模型通过SPARQL式查询对结构化知识图谱进行推理，可以提升知识支撑的推理效率和可控性；这是改进我们流程中可靠营养知识检索的一条有前景的方向 [27]。

参考文献

[1] J.-S. Shim、K. Oh 和 H. C. Kim，《流行病学研究中的膳食评估方法》，*Epidemiology and Health*，第 36 卷，2014009，2014 年，第 e 页。[2] M. Kay、J. Santos 和 M. Takane，《移动健康：通过移动技术开拓健康新视野》，*World Health Organization*，第 64 卷，第 7 期，第 66–71 页，2011 年。[3] L. M. K¨onig、M. Van Emmenis、J. Nurmi、A. Kassavou 和 S. Sutton，《基于智能手机的膳食评估工具的特征：系统综述》，*Health Psychology Review*，第 16 卷，第 4 期，第 526–550 页，2022 年。[4] C.-S. Lee、M.-H. Wang、G. Acampora、C.-Y. Hsu 和 H. Hagras，《基于二型模糊本体和模糊标记语言的膳食评估》，*International Journal of Intelligent Systems*，第 25 卷，第 12 期，第 1187–1216 页，2010 年。[5] D. Tsolakidis、L. P. Gymnopoulos 和 K. Dimitropoulos，《用于个性化营养的人工智能和机器学习技术：综述》，载于*Proc. Informatics*，第 11 卷，第 3 期，第 62 页，2024 年。[6] T. I vas¸cu、A. Dinis¸ 和 K. Cincar，《基于多智能体架构的疾病驱动型营养推荐系统》，载于*Proc. 8th Int. Conf. Web Intelligence, Mining and Semantics (WIMS)*，2018 年，第 1–5 页。

[7] S. Iqbal、W. Altaf、M. Aslam、W. Mahmood 和 M. U. G. Khan，《智能体在医疗保健中的应用》，Artificial Intelligence Review，第 46 卷，第 1 期，第 83–112 页，2016 年。[8] 美国国家科学院、工程院和医学院，《钠和钾的膳食参考摄入量》，2019 年。[9] H. B. Del Valle、A. L. Yaktine、C. L. Taylor 和 A. C. Ross，《钙和维生素D的膳食参考摄入量》，美国国家科学院出版社，2011 年。[10] 膳食参考摄入量科学评估常设委员会和电解质与水膳食参考摄入量专家组，《水、钾、钠、氯化物和硫酸盐的膳食参考摄入量》。美国国家科学院出版社，2005 年。[11] E. Chin、Y. Bouzid、D. Lemay、J. Smilowitz 和 Y. Vainberg，《餐食照片营养素评估调查研究方案（SNAPMe）：食物记录及配对餐食照片的基准数据集》，Current Developments in Nutrition，第 6 卷，第 351 页，2022 年。[12] W. Min、Z. Wang、Y. Liu、M. Luo、L. Kang、X. Wei 等，《大规模视觉食物识别》，IEEE Transactions on Pattern Analysis and Machine Intelligence，第 45 卷，第 8 期，第 9932–9949 页，2023 年 8 月。[13] J. Mar´ın、A. Biswas、F. Ofli、N. Hynes、A. Salvador、Y. Aytar 等，《Recipe1M+：用于学习烹饪食谱与食物图像跨模态嵌入的数据集》，IEEE Transactions on Pattern Analysis and Machine Intelligence，第 43 卷，第 1 期，第 187–203 页，2021 年 1 月。[14] T. Perrett、A. Darkhalil、S. Sinha、O. Emara、S. Pollard、K. K. Parida 等，《Hd-epic：一个高度细致的第一人称视角视频数据集》，载于 Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)，2025 年，第 23901–23913 页。[15] S. Jiang、W. Min、L. Liu 和 Z. Luo，《用于食物识别的多尺度多视图深度特征聚合》，IEEE Transactions on Image Processing，第 29 卷，第 265–276 页，2020 年。[16] W. Min、L. Liu、Z. Luo 和 S. Jiang，《用于食物识别的成分引导级联多注意力网络》，载于 Proc. 27th ACM Int. Conf. Multimedia (ACM MM)，2019 年，第 1331–1339 页。[17] T. Perrett、A. Darkhalil、S. Sinha、O. Emara、S. Pollard、K. K. Parida 等，《Hd-epic：一个高度细致的第一人称视角视频数据集》，载于 Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)，2025 年，第 23901–23913 页。[18] F. P.-W. Lo、Y. Sun、J. Qiu 和 B. P. L. Lo，《Point2Volume：一种利用视图合成的基于视觉的膳食评估方法》，IEEE Transactions on Industrial Informatics，第 16 卷，第 1 期，第 577–586 页，2020 年 1 月。[19] A. Salvador、E. Gundogdu、L. Bazzani 和 M. Donoser，《利用层次化Transformer和自监督学习改进跨模态食谱检索》，载于 Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)，2021 年，第 15470–15479 页。[20] H. Wang、G. Lin、S. C. H. Hoi 和 C. Miao，《用于从图像生成食谱的结构感知生成网络》，载于 Proc. Eur. Conf. Comput. Vis. (ECCV)，2020 年，第 359–374 页。[21] A. Saphala、R. Zhang 和 O. Amft，《利用专家和数据模型在智能眼镜上进行基于接近度的进食事件检测》，载于 Proc. ACM Int. Symp. Wearable Computers (ISWC)，2022 年，第 59–63 页。[22] K. Kyritsis、C. Diou 和 A. Delopoulos，《一种利用智能手表在真实环境中监测进食行为的数据驱动端到端方法》，IEEE Journal of Biomedical and Health Informatics，第 25 卷，第 1 期，第 22–34 页，2021 年 1 月。[23] F. P.-W. Lo、Y. Sun、J. Qiu 和 B. Lo，《用于膳食评估的基于图像的食物分类与体积估算：综述》，IEEE Journal of Biomedical and Health Informatics，第 24 卷，第 7 期，第 1926–1939 页，2020 年 7 月。[24] F. P.-W. Lo、J. Qiu、Z. Wang、J. Chen、B. Xiao、W. Yuan 等，《使用多模态ChatGPT进行膳食评估：系统分析》，IEEE Journal of Biomedical and Health Informatics，即将刊出，2024 年。[25] Z. Wang、Y. Yu、W. Zheng、W. Ma 和 M. Zhang，《Macrec：一种用于推荐的多智能体协作框架》，载于 Proc. 47th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. (SIGIR)，2024 年，第 2760–2764 页。[26] P. Anderson、A. X. Chang、D. S. Chaplot、A. Dosovitskiy、S. Gupta、V. Koltun 等，《关于具身导航智能体的评估》，arXiv 预印本 arXiv:1807.06757，2018 年。[27] M. Lan、Y. Xia、G. Zhou、N. Huang、Z. Li 和 H. Wu，《LLM4QA：利用大语言模型，通过SPARQL查询高效进行知识图谱推理》，Journal of Advances in Information Technology，第 15 卷，第 10 期，第 1157–1162 页，2024 年。