

团 体 标 准

T/ISC XXXX—XXXX

基于医疗健康画像的大模型能力效果评估 方法

Method for evaluating the capability and effect of large models based on
healthcare portrait

(征求意见稿)

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX - XX - XX 发布

XXXX - XX - XX 实施

中 国 互 联 网 协 会 发 布

目 次

| | |
|------------------------------|----|
| 前 言 | II |
| 1 范围 | 3 |
| 2 规范性引用文件 | 3 |
| 3 术语和定义 | 3 |
| 4 符号和缩略语 | 3 |
| 5 基于健康画像的大模型能力通用效果评估 | 3 |
| 5.1 评估对象 | 3 |
| 5.2 通用评估流程 | 3 |
| 5.3 通用评估方法 | 4 |
| 5.4 通用评估指标 | 4 |
| 5.5 数据集通用要求 | 6 |
| 6 基于健康画像的大模型能力典型场景效果评估 | 6 |
| 6.1 症状咨询场景评估 | 6 |
| 6.2 用药咨询场景评估 | 7 |
| 6.3 检查检验报告解读场景评估 | 7 |
| 6.4 智能导诊与分诊场景评估 | 8 |
| 6.5 辅助诊疗场景评估 | 8 |
| 6.6 疾病风险预测场景评估 | 9 |
| 6.7 病情评估场景评估 | 9 |
| 6.8 医嘱质控场景评估 | 10 |
| 6.9 慢病诊后管理场景评估 | 11 |
| 6.10 饮食运动建议场景评估 | 11 |

前　　言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国互联网协会提出并归口。

本文件起草单位：

本文件主要起草人：

本文件及其所代替文件的历次版本发布情况为：

——

基于医疗健康画像的大模型能力效果评估方法

1 范围

本文件规定了基于医疗健康画像的大模型能力效果评价及对应的评价指标要求，明确了健康画像增强生成的大模型能力效果的评估方向与核心维度。

本标准适用于基于医疗健康画像的大模型能力效果评估活动，覆盖医疗机构、医疗科技企业、医疗健康数据服务机构、公共卫生管理等部门等相关单位，可作为医疗健康大模型在健康咨询、辅助诊疗、公共卫生决策等应用效果的评估依据之一。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

WS/T 363-2023 卫生健康信息数据元目录

WS/T 364-2023 卫生健康信息数据元值域代码

WS/T 846-2024 医院信息平台交互标准

IEEE P3394 大语言模型智能体界面标准（Standard for Large Language Model Agent Interface）

ITU-T F.748.44 基础模型评估标准（A Assessment criteria for foundation models – Benchmark）

3 术语和定义

3.1

医疗健康画像 healthcare portrait

是指将居民个人多维度健康数据进行采集、清洗、整合与分析，形成的全方位、数字化、标签化的个人健康模型。

4 符号和缩略语

下列符号和缩略语适用于本文件。

AUC: 曲线下面积（Area Under the Curve）

FN: 假阴性（False Negative）

FP: 假阳性（False Positive）

ICD-10: 国际疾病分类第十版（International Classification of Disease, 10th Revision）

ROC: 接受者操作特性曲线（Receiver Operating Characteristic）

SNOMED CT: 医学系统命名法——临床术语（Systematized Nomenclature of Medicine – Clinical Terms）

TN: 真阴性（True Negative）

TP: 真阳性（True Positive）

5 基于健康画像的大模型能力通用效果评估

5.1 评估对象

通用评估流程与通用评估指标适用于各类应用健康画像增强生成的大模型的能力评估，针对不同场景的大模型应用基于场景设置个性化评估指标。

5.2 通用评估流程

遵循“准备-实施-分析-优化-总结”的逻辑，结合医疗健康画像的特殊性与数据合规要求，计算各场景对应指标，具体流程如下：

- a) 评估准备：明确评估目标、评估范围及评估对象，确定评估所覆盖的医疗场景，制定评估方案，明确数据来源、评估指标、评估方法及时间节点，同步完成评估工具的调试与校准；
- b) 数据采集与预处理：采集符合评估要求的医疗健康画像数据及对应场景的模型输入输出数据，数据来源需符合医疗数据合规要求，涵盖不同人群、不同场景的样本，确保样本的代表性与多样性；对采集的数据进行预处理，生成评估数据集；
- c) 模型部署与测试：输入基础数据+对应健康画像数据，模拟有健康画像支撑的大模型应用场景，记录大模型输出结果、响应速度等实时数据；
- d) 指标计算：根据模型测试记录的输入输出数据，按照评估指标计算公式，保存测试日志，确保指标计算的准确性与可追溯性；
- e) 结果分析与优化：对比评估指标的临床需求标准，分析各个场景下的指标与临床需求标准的差距；
- f) 评估报告生成：汇总评估过程、数据来源、指标计算结果、结果分析及优化建议，形成标准化评估报告。

5.3 通用评估方法

采用自动化与人工结合的多维度评估体系，具体方法如下：

- a) 数据抽样方法：采用分层抽样法选取评估样本，根据医疗健康画像的核心维度与场景类型进行分层。确保每层样本数量占比与实际应用场景中的分布比例一致，确保样本覆盖全面。记录抽样规则、样本数量及分层明细；
- b) 模型测试方法：采用批量数据输入模式，基于预处理后的测试集自动化运行模型，记录输出结果；
- c) 指标计算方法：采用“自动化工具+手动校验”结合的方式。自动化工具需符合医疗数据规范，实现指标的批量计算。手动校验需选取不少于10%的测试样本，由多位领域专家判断自动化工具输出指标是否正确，确保自动化测试与人工判断的一致性。对于纯粹的分类任务，直接使用自动化工具计算指标。对于生成类等没有标准答案的测试任务，在计算指标的基础上，需抽样进行人工评价；

5.4 通用评估指标

5.4.1 分类任务技术指标

本类指标适用于所有判别性任务（如命名实体识别、异常项判定等），即模型将输入数据（如文本、图像等）划分到预定类别中的能力：

- d) 准确率：针对所有分类任务，计算模型所有预测中正确预测的总体比例，计算公式如下：

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

式中：

Accuracy——准确率；

TP——真阳性数量；

TN——真阴性数量；

FP——假阳性数量；

FN——假阴性数量。

- e) 精确率：针对二分类任务，计算模型预测为正例的结果中，实际为正例的比例，计算公式如下：

$$Precision = \frac{TP}{TP+FP}$$

式中：

Precision——精确率；

TP——真阳性数量；

FP——假阳性数量。

f) 召回率：针对二分类任务，计算所有实际为正例的样本中，被模型正确识别为正例的比例，计算公式如下：

$$Recall = \frac{TP}{TP+FN}$$

式中：

Recall——召回率；

TP——真阳性数量；

FN——假阴性数量。

g) F1 分数：针对二分类任务，计算精确率与召回率的调和平均数，计算公式如下：

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

式中：

F1——F1分数；

Precision——精确率；

Recall——召回率。

h) ROC-AUC：针对二分类任务，计算 ROC 曲线下的面积 AUC，计算公式如下：

$$AUC = \int_0^1 TPR \, dFPR$$

式中：

TPR——真阳性率；

FPR——假阳性率。

i) 宏精确率：针对多分类任务，计算每个类别的精确率再求算数平均，计算公式如下：

$$Macro - P = \frac{1}{K} \sum_{i=1}^K Precision_i$$

式中：

Macro-P——宏精确率；

K——分类任务总类别数；

Precision_i——第i类精准率。

j) 宏召回率：针对多分类任务，计算每个类别的召回率再求算数平均，计算公式如下：

$$Macro - R = \frac{1}{K} \sum_{i=1}^K Recall_i$$

式中：

Macro-R——宏召回率；

K——分类任务总类别数；

Recall_i——第i类召回率。

k) 宏 F1 分数：针对多分类任务，计算宏精确率与宏召回率的调和平均数，计算公式如下：

$$Macro - F1 = \frac{2 \times Macro - P \times Macro - R}{Macro - P + Macro - R}$$

式中：

Macro-F1——宏F1分数；

Macro-P——宏精确率；

Macro-R——宏召回率。

5.4.2 生成类任务技术指标

本类指标适用于所有自然语言生成任务，即模型根据结构化数据或非结构化数据输入自动生成连贯文本的能力：

1) BERTScore：对生成任务，计算客观指标 BERTScore，计算公式如下：

$$sim(x_i, y_i) = \frac{Emb(x_i) \cdot Emb(y_i)}{\|Emb(x_i)\| \|Emb(y_i)\|}$$

$$Precision = \frac{1}{|x|} \sum_{x_i} \max_{y_i \in y} sim(x_i, y_i)$$

$$Recall = \frac{1}{|y|} \sum_{y_i} \max_{x_i \in x} sim(x_i, y_i)$$

$$BERTScore = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

式中：

$Emb(x_i)$ ——句子x中词语在经过编码器后的嵌入向量；

$Emb(y_i)$ ——句子y中词语在经过编码器后的嵌入向量；

$sim(x_i, y_i)$ ——两词语嵌入向量的余弦相似度；

Precision——精确率；

Recall——召回率。

m) ROUGE-N：对生成任务，计算客观指标 ROUGE-N，其计算公式如下：

$$ROUGE - N = \frac{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} count_{match}(gram_n)}{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} count(gram_n)}$$

式中：

N ——即n-gram，文本内容滑动窗口字节数，参考值为2；

$Count_{match}(gram_n)$ ——参考摘要和生成摘要中共有的n-gram的数量；

$Count(gram_n)$ ——参考摘要中n-gram的数量。

5.5 数据集通用要求

基于医疗健康画像的大模型能力效果评估数据集应包括但不限于下列字段：

- a) id: 每个数据集的标识符；
- b) scenario_name: 问题所属场景；
- c) type: 问题类型（如二分类、多分类、简答题等）；
- d) question: 问题文本；
- e) answer: 参考答案；
- f) exp: 对答案的解释。

6 基于健康画像的大模型能力典型场景效果评估

6.1 症状咨询场景评估

6.1.1 功能要求

基于医疗健康画像，大模型应支持对患者输入的症状描述，咨询需求进行精准识别与意图判断，结合患者既往病史、过敏史、生理指标等画像信息，明确咨询核心需求并生成贴合个体情况的咨询回复：

- a) 应支持从患者非结构化症状描述（如“最近三天发烧、咳嗽，伴有乏力”）中，识别核心症状、症状持续时间、伴随症状等关键信息，结合患者健康画像中的年龄、基础疾病等维度，区分症状的轻重程度；
- b) 应支持准确识别患者的咨询意图（如症状原因咨询、缓解方法咨询、就医建议咨询等），区分正常咨询与紧急情况（如剧烈疼痛、呼吸困难等），并给出对应提示；
- c) 应支持识别患者对症状的否定描述（如“无发烧症状”）与不确定描述（如“疑似咳嗽”），结合健康画像历史数据进行交叉验证，提升意图理解的准确性；
- d) 应支持结合患者基础疾病、过敏史、用药情况等画像信息，生成针对性的症状缓解建议，避免推荐与患者健康状况冲突的方法；
- e) 应支持将咨询回复结构化呈现，清晰说明症状可能原因、缓解方法、就医提示等内容，语言通俗易懂，符合患者认知水平；
- f) 应支持基于患者历史咨询记录与健康画像变化，进行纵向分析，为患者提供连贯的咨询服务。

6.1.2 准确度要求

本功能涉及指标如下，通用技术指标均应符合第5章要求：

- a) 准确率（应符合 5.1.1 计算公式）；
- b) 精确率（应符合 5.1.1 计算公式）；
- c) 召回率（应符合 5.1.1 计算公式）；
- d) F1 分数（应符合 5.1.1 计算公式）；
- e) BERTScore（应符合 5.1.2 计算公式）；
- f) ROUGE-N（应符合 5.1.2 计算公式）。

6.2 用药咨询场景评估

6.2.1 功能要求

基于医疗健康画像，大模型应支持对患者咨询的药物名称、用药相关问题进行精准识别，结合患者画像信息解读用药要点并给出个性化用药指导：

- a) 应支持识别患者咨询的药物名称、用药剂量、用药频次、用药时长等核心信息，准确区分相似药物（如名称相近、剂型不同的药物）；
- b) 应支持结合患者健康画像中的检验结果、过敏史、基础疾病、正在使用的其他药物等信息，解读药物的适用性、禁忌证、注意事项；
- c) 应支持识别患者对用药问题的疑问点，针对性给出清晰解答；
- d) 应支持根据患者年龄、体重、肝肾功能等画像维度，调整用药建议，如老年人、儿童的剂量调整提示；
- e) 应支持将用药指导结构化呈现，明确药物用法用量、禁忌人群、注意事项、药物相互作用等内容，语言通俗易懂，避免专业术语过于晦涩；
- f) 应支持基于患者用药历史与健康画像变化，提醒患者按时用药、调整用药方案，提供连贯的用药咨询服务。

6.2.2 准确度要求

本功能涉及指标如下，通用技术指标均应符合第5章要求：

- a) 准确率（应符合 5.1.1 计算公式）；
- b) 精确率（应符合 5.1.1 计算公式）；
- c) 召回率（应符合 5.1.1 计算公式）；
- d) F1 分数（应符合 5.1.1 计算公式）；
- e) BERTScore（应符合 5.1.2 计算公式）；
- f) ROUGE-N（应符合 5.1.2 计算公式）。

6.3 检查检验报告解读场景评估

6.3.1 功能要求

基于医疗健康画像，大模型应支持对检查检验报告中的各项指标、结果进行精准提取，结合患者历史画像进行趋势分析与解读：

- a) 应支持从检查检验报告（如血常规、尿常规、影像学报告等）中，提取各项检测指标名称、检测数值、参考范围、结果判定（如正常、异常升高、异常降低）等核心信息；
- b) 应支持结合患者健康画像中的历史检查检验数据，构建个体历史基线，对当前检查进行联合性、对比性解读，体现指标变化趋势，而非仅复述数值；
- c) 应支持识别异常检测结果，结合患者基础疾病、症状等画像信息，解读异常结果的可能原因、临床意义，避免夸大或遗漏异常风险；
- d) 应支持将解读结果按检测项目分类呈现，清晰标注异常指标、趋势变化、可能原因及就医提示；
- e) 应支持结合患者健康画像中的基础疾病，针对异常结果给出针对性的干预建议（如饮食调整、进一步检查提示）；
- f) 应支持将解读结果与标准医学语体系（如 ICD-10、SNOMED CT）对接，便于医师结合临床场景参考使用。

6.3.2 准确度要求

本功能涉及指标如下，通用技术指标均应符合第5章要求：

- a) 准确率（应符合 5.1.1 计算公式）；
- b) 精确率（应符合 5.1.1 计算公式）；
- c) 召回率（应符合 5.1.1 计算公式）；
- d) F1 分数（应符合 5.1.1 计算公式）；
- e) BERTScore（应符合 5.1.2 计算公式）；
- f) ROUGE-N（应符合 5.1.2 计算公式）。

6.4 智能导诊与分诊场景评估

6.4.1 功能要求

主诉与症状识别：基于医疗健康画像，大模型应支持对患者主诉、症状描述进行精准识别，结合患者历史画像信息，明确分诊核心依据：

- a) 应支持从患者主诉、症状描述中，提取核心症状、症状持续时间、严重程度等关键信息，结合患者健康画像中的年龄、基础疾病、过敏史等维度，综合判断病情紧急程度；
- b) 应支持识别患者描述中的模糊信息（如“身体不舒服”等），通过简单交互补充关键信息，提升导诊/分诊的准确性；
- c) 应支持结合健康画像，了解患者既往就诊科室、疾病诊断，辅助优化导诊/分诊建议；
- d) 应支持根据患者症状、病情紧急程度及健康画像信息，推荐最适配的就诊科室（如发烧咳嗽推荐呼吸内科、关节疼痛推荐骨科），避免错分、误分；
- e) 应支持区分普通就诊与急诊，对紧急情况（如胸痛、呕血）优先推荐急诊科，并给出紧急就医提示；
- f) 应支持将导诊/分诊建议清晰呈现，明确推荐科室、就诊注意事项（如携带既往检查报告），辅助患者快速就诊。

6.4.2 准确度要求

本功能涉及指标如下，通用技术指标均应符合第5章要求：

- a) 准确率（应符合 5.1.1 计算公式）；
- b) 精确率（应符合 5.1.1 计算公式）；
- c) 召回率（应符合 5.1.1 计算公式）；
- d) F1 分数（应符合 5.1.1 计算公式）；
- e) BERTScore（应符合 5.1.2 计算公式）；
- f) ROUGE-N（应符合 5.1.2 计算公式）。

6.5 辅助诊疗场景评估

6.5.1 功能要求

诊疗信息整合与分析：基于医疗健康画像，大模型应支持整合患者主诉、症状、检查检验结果、既往病史等多维度信息，为医师提供辅助诊疗参考：

- a) 应支持从患者健康画像中，整合现病史、既往史、个人史、家族史、检查检验结果、用药史等核心诊疗信息，进行结构化整理，便于医师快速查看；
- b) 应支持结合患者症状与检查检验结果，分析可能的疾病诊断方向，结合医学规范给出诊断参考建议；
- c) 应支持针对可能的诊断方向，推荐合适的治疗方案（如药物治疗、手术治疗、康复治疗），结合患者肝肾功能、过敏史等画像信息，排除不匹配的治疗方案；
- d) 应支持将诊断参考、治疗推荐按优先级排序，清晰标注推荐依据（如结合患者血常规异常、咳嗽症状，推荐呼吸内科相关诊断）；
- e) 应支持医师补充诊疗信息（如新增检查结果）后，模型实时调整诊断与治疗推荐建议；
- f) 应支持将诊疗建议与标准医学术语体系、临床诊疗指南对接，确保建议的规范性与专业性。

6.5.2 准确度要求

本功能涉及指标如下，通用技术指标均应符合第5章要求：

- a) 准确率（应符合 5.1.1 计算公式）；
- b) 精确率（应符合 5.1.1 计算公式）；
- c) 召回率（应符合 5.1.1 计算公式）；
- d) F1 分数（应符合 5.1.1 计算公式）；
- e) BERTScore（应符合 5.1.2 计算公式）；
- f) ROUGE-N（应符合 5.1.2 计算公式）。
- g) 召回率@k：对于诊断结果中，包含真实疾病的比列，计算公式如下：

$$Recall@k = \frac{|R_k \cap T|}{|T|}$$

式中：

T——真实相关集合；

R_k——模型预测的前K个结果集合。

6.6 疾病风险预测场景评估

6.6.1 功能要求

风险因素提取与分析：基于医疗健康画像，大模型应支持提取患者与疾病相关的风险因素，结合画像数据进行风险分析：

- a) 应支持从患者健康画像中，提取与目标疾病相关的风险因素，包括基因数据、生理指标、生活习惯、既往病史、家族病史等核心维度；
- b) 应支持针对不同类型的疾病（如慢病、传染病），结合对应的风险因素，构建风险评估模型，分析患者发病风险等级（如低风险、中风险、高风险）；
- c) 应支持结合患者健康画像的动态变化（如生理指标变化、生活习惯调整），实时更新风险因素分析结果；
- d) 应支持基于患者风险因素，预测未来一定周期内（如 1 年、3 年）的发病风险，明确风险等级与预测依据；
- e) 应支持针对高风险患者，结合其健康画像，给出针对性的风险干预建议（如饮食调整、运动建议、定期检查提示）；
- f) 应支持对传染病等公共卫生相关疾病，结合区域人群画像与个体画像，给出风险预警提示，辅助公共卫生管理。

6.6.2 准确度要求

本功能涉及指标如下，通用技术指标均应符合第5章要求：

- a) 准确率（应符合 5.1.1 计算公式）；
- b) 精确率（应符合 5.1.1 计算公式）；
- c) 召回率（应符合 5.1.1 计算公式）；
- d) F1 分数（应符合 5.1.1 计算公式）；
- e) BERTScore（应符合 5.1.2 计算公式）；
- f) ROUGE-N（应符合 5.1.2 计算公式）。
- g) C 指数：即 C-index，指在所有可比较样本对中，预测风险与实际事件一致样本对所占的比例，计算公式如下：

$$C = \frac{\sum_{i,j} I(T_i < T_j \cap \eta_i > \eta_j)}{\sum_{i,j} I(T_i < T_j)}$$

式中：

T_i——个体i的实际观察时间；

T_j——个体j的实际观察时间；

η_i——个体i的预测风险分数；

η_j——个体j的预测风险分数。

6.7 病情评估场景评估

6.7.1 功能要求

基于医疗健康画像，大模型应支持整合患者当前病情、治疗情况、既往病史等多维度信息，为病情评估提供全面依据：

- a) 应支持从患者健康画像中，提取当前症状、检查检验结果、用药情况、治疗效果等核心病情信息，结合既往病史、基础疾病，构建完整的病情评估数据集；
- b) 应支持识别患者病情的动态变化（如症状缓解、加重，指标改善、恶化），结合治疗方案调整情况，分析病情变化原因与恢复情况；
- c) 应支持结合患者年龄、身体状况、心理状态等画像维度，综合考量病情严重程度与恢复潜力。
- d) 应支持对患者当前病情进行分级评估（如轻度、中度、重度），明确评估依据，客观反映病情严重程度；
- e) 应支持结合病情评估结果，给出针对性的治疗调整建议、护理建议（如饮食护理、康复训练），贴合患者个体情况；
- f) 应支持将病情评估结果结构化呈现，清晰标注评估结论、病情变化趋势、后续建议，便于医师与患者参考。

6.7.2 准确度要求

本功能涉及指标如下，通用技术指标均应符合第5章要求：

- a) 准确率（应符合 5.1.1 计算公式）；
- b) 精确率（应符合 5.1.1 计算公式）；
- c) 召回率（应符合 5.1.1 计算公式）；
- d) F1 分数（应符合 5.1.1 计算公式）；
- e) BERTScore（应符合 5.1.2 计算公式）；
- f) ROUGE-N（应符合 5.1.2 计算公式）。

6.8 医嘱质控场景评估

6.8.1 功能要求

基于医疗健康画像，大模型应支持对医师开具的医嘱进行精准提取与解析，结合患者画像信息识别不合理医嘱：

- a) 应支持从医嘱中提取用药医嘱、检查检验医嘱、护理医嘱等核心信息，准确识别药物名称、剂量、频次、检查项目等关键内容；
- b) 应支持结合患者健康画像中的过敏史、基础疾病、正在使用的其他药物、检查检验结果等信息，解析医嘱的合理性；
- c) 应支持识别各类不合理医嘱，包括用药禁忌、剂量异常、药物相互作用、检查检验项目重复或不适当配等情况。
- d) 应支持在医师开具医嘱后，实时检测不合理项，及时发出拦截提示，明确标注不合理原因（如“患者对青霉素过敏，医嘱开具青霉素类药物”）；
- e) 应支持结合患者健康画像与临床规范，针对不合理医嘱给出合理的修正建议（如替换适配药物、调整剂量）；
- f) 应支持记录医嘱质控结果，包括合理医嘱数量、不合理医嘱数量、不合理类型等，为后续质控优化提供依据。

6.8.2 准确度要求

本功能涉及指标如下，通用技术指标均应符合第5章要求：

- a) 准确率（应符合 5.1.1 计算公式）；
- b) 精确率（应符合 5.1.1 计算公式）；
- c) 召回率（应符合 5.1.1 计算公式）；
- d) F1 分数（应符合 5.1.1 计算公式）；
- e) BERTScore（应符合 5.1.2 计算公式）；
- f) ROUGE-N（应符合 5.1.2 计算公式）。

6.9 慢病诊后管理场景评估

6.9.1 功能要求

基于医疗健康画像，大模型应支持对慢性病（如高血压、糖尿病等）患者诊后的长期、连续、个性化管理，整合监测数据、行为习惯与治疗方案，提供动态的管理支持与干预建议：

- a) 应支持从患者健康画像中，持续跟踪与管理目标慢性病相关的核心指标（如血压值、血糖值、血脂水平等），并与预设的管理目标值进行对比分析；
- b) 应支持整合患者诊后的用药记录、症状变化、生活方式数据（如饮食、运动、睡眠），评估治疗方案的执行效果与依从性；
- c) 应支持识别慢性病管理的异常情况，如指标控制不佳、症状反复、用药中断等，结合画像信息分析潜在原因（如饮食不当、运动不足、药物副作用等）；
- d) 应支持基于患者的近期指标趋势、行为数据及画像中的并发症风险，生成个性化的阶段性管理计划调整建议，包括用药调整、监测频率优化、生活方式干预强化等；
- e) 应支持提供定期的管理总结与教育内容，以通俗易懂的方式向患者反馈阶段管理成果、解释指标意义、强化健康知识，提升患者自我管理能力；
- f) 应支持设置关键指标预警阈值，在患者数据（如手动录入或设备上传的指标）触发预警时，及时生成提醒或建议，必要时提示就医。

6.9.2 准确度要求

本功能涉及指标如下，通用技术指标均应符合第5章要求：

- a) 准确率（应符合 5.1.1 计算公式）；
- b) 精确率（应符合 5.1.1 计算公式）；
- c) 召回率（应符合 5.1.1 计算公式）；
- d) F1 分数（应符合 5.1.1 计算公式）；
- e) BERTScore（应符合 5.1.2 计算公式）；
- f) ROUGE-N（应符合 5.1.2 计算公式）。

6.10 饮食运动建议场景评估

6.10.1 功能要求

基于医疗健康画像，大模型应支持根据患者的疾病状况、生理指标、营养需求、运动能力及个人偏好，生成个性化、安全、可执行的饮食与运动建议：

- a) 应支持从患者健康画像中，准确提取与饮食运动建议相关的关键信息，包括疾病诊断（如糖尿病、肾病）、身体质量指数（BMI）、食物过敏或食物不耐受、检验检查指标（如血糖、尿酸）、运动禁忌、体能水平、个人饮食偏好与地域文化习惯等；
- b) 应支持结合患者的疾病管理目标（如减重、控糖、降压、增肌），计算推荐每日能量及营养素摄入范围，并基于此生成具体的饮食建议，包括食物种类、份量、餐次分配及食谱示例；
- c) 应支持结合患者的身体状况、运动风险及兴趣，推荐适宜的运动类型、强度、频率、时长及注意事项，避免推荐存在安全隐患的运动方式；
- d) 应支持识别饮食或运动建议与患者健康画像的潜在冲突，如为肾病患者推荐高钾食物、为关节损伤者推荐高强度冲击性运动等，并予以规避或明确警示；
- e) 应支持将饮食运动建议结构化、分点呈现，内容具体、量化（如“每日主食摄入约 200-250 克，粗细搭配”），语言鼓励性强，易于患者理解和执行；
- f) 应支持在患者提供近期执行反馈（如饮食日志、运动记录）后，能对建议进行适应性微调，形成动态优化的健康生活方案。

6.10.2 准确度要求

本功能涉及指标如下，通用技术指标均应符合第5章要求：

- a) 准确率（应符合 5.1.1 计算公式）；
- b) 精确率（应符合 5.1.1 计算公式）；
- c) 召回率（应符合 5.1.1 计算公式）；

- d) F1 分数（应符合 5.1.1 计算公式）；
 - e) BERTScore（应符合 5.1.2 计算公式）；
 - f) ROUGE-N（应符合 5.1.2 计算公式）。
-