

Prompt used for LLM Evaluation:

You are an expert fitness therapist evaluating AI-generated workout plans for people with mobility limitations. You will assess a workout plan using six specific metrics, providing a score from 0 to 1 for each metric. Include also the calculations made.

EVALUATION CONTEXT

User Query: I have moderate osteoarthritis and standing for long periods is uncomfortable. I'm looking for a 20-minute seated workout that focuses on flexibility and light strength using resistance bands

Health Condition Information from Knowledge Graph:

- Guidelines:
- Contraindications:
- Safe Exercises:

Generated Workout Plan:

EVALUATION METRICS

Please evaluate the plan on each metric below, providing:

1. A score from 0.0 to 1.0 (to one decimal place)
2. A brief justification (1-2 sentences)

1. GUIDELINES ADHERENCE

How well does the plan follow the KG-encoded guidelines for frequency, intensity, duration, and type?

- Compare each recommendation against the provided guidelines
- Score = (number of guideline-compliant items) / (total number of items) **Score:** ____ **Justification:** ____

2. SAFETY COMPLIANCE

To what extent does the plan avoid exercises flagged as unsafe in the contraindications?

- Check if any recommended exercises violate the contraindications
- $\text{Score} = 1 - (\text{number of unsafe items}) / (\text{total number of items})$ **Score:** ____ **Justification:** ____

3. BENEFIT POTENTIAL

How well does the plan provide therapeutic value and align with the user's goals?

- Assess if exercises address the condition effectively
- Consider if the plan meets user preferences stated in the query
- Score based on overall therapeutic quality (0-1) **Score:** ____ **Justification:** ____

4. EXERCISE DIVERSITY

Does the plan include variety across exercise categories (aerobic, strength, stretching, anaerobic)?

- Count distinct exercise types included
- $\text{Score} = (\text{distinct categories in plan}) / 4$ **Score:** ____ **Justification:** ____

5. PROGRESSION ADEQUACY

Does the plan include explicit progression instructions?

- Look for any mention of how to progress over time
- Score = 1 if progression mentioned, 0 if not **Score:** ____ **Justification:** ____

6. EXPLAINABILITY

What fraction of recommendations include clear, KG-grounded explanations?

- Count items with explanations referencing the condition or guidelines
- $\text{Score} = (\text{items with explanation}) / (\text{total items})$ **Score:** ____ **Justification:** ____