#### Pointwise-to-Pairwise

- 1. Combined with the reference and LLM-generated texts with their pointwise grading results, give a detailed comparison critique ... (*Knowledge Injection*)
- 2. Give a detailed explanation if the comparison label isn't consistent with the rating order ... (Self-Reflection)

Reference Text

# Pointwise Grading

Evaluation Input

User Query

LLM-generated Text

Evaluation Output (Critique)

**Explanation**: Compared with the reference, AI assistant's response ...

Rating: [5]

Referenced

Path#2

Path#1

 $f_{P2P}$ 

 $f_{R2RF}$ 

### Referenced-to-Reference-Free

- 1. Don't mention the comparison with references directly, while keeping enough details ... (*Reference Removal*)
- 2. Allow to modify the rating if the rating isn't consistent with the revised explanation ... (Self-Reflection)

# Pairwise Comparison

#### **Evaluation Input**

User Query ) ( LLM-generated Text 1

Reference Text LLM-generated Text 2

#### **Evaluation Output (Critique)**

**Explanation**: Based the reference, Al assistant 1's response ..., while Al assistant 2's response ... **Label**: Al assistant 1 Wins

Referenced

## Pointwise Grading

**Evaluation Input** 

User Query

LLM-generated Text

**Evaluation Output (Critique)** 

**Explanation**: Al assistant's response ...

Rating: [5]

## Reference-Free

### Referenced-to-Reference-Free

Don't mention the comparison with references directly, while keeping enough details ... (*Reference Removal*)
Allow to slightly modify the comparison label if the label isn't consistent with the revised explanation ... (*Self-Reflection*)

 $f_{R2RF}$ 

Cross

**Validation** 

 $f_{P2P}$ 

## Pairwise Comparison

**Evaluation Input** 

**User Query** 

LLM-generated Text 2

LLM-generated Text 1

**Evaluation Output (Critique)** 

Explanation: Al assistant 1's response ..., while Al assistant 2's response ...

Label: Al assistant 1 Wins

Reference-Free

#### Pointwise-to-Pairwise

- 1. Combined with LLM-generated texts with their pointwise grading results, give a detailed comparison critique ... (*Knowledge Injection*)
- 2. Give a detailed explanation if the comparison label isn't consistent with the rating order ... (Self-Reflection)