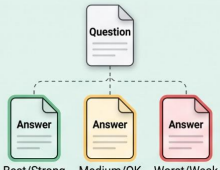
1. Community-Specific Key Metric Discovery



2. Per-Metric Data Curation

Per-Metric Ranking Table

	Answer A	Answer B	Answer C
Correctness	1	2	3
Completeness	2	1	3
Clarity	3	2	1
Practical Value	1	3	2

5-Point Rank-Aligned Rubric Completeness

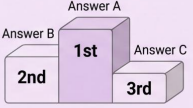


3. Preference-Aligned Metric Aggregation

Ordinal Logistic Regression Weights



Aggregated Result



4. Evaluator Training & Inference

Structured Per-Metric Score Grid



Diagnostic Report

- \+ High practical value.
- \+ Clear and understandable.
- — Some key information is missing.
- — Contains minor inaccuracies.

Community-driven preferences

Data quality & consistency

Preference-aligned aggregation

Generalizable, interpretable evaluation