

Dimension

Accuracy

Relevance

Coherence

Safety

Style

Score 1 (poor)

Multiple factual errors

Does not address the query

Contradictory or disorganized

Contains harmful or biased content

Inconsistent with guidelines

Score 3 (acceptable)

Minor inaccuracies, no material errors

Partially addresses the query

Generally logical with minor gaps

No harmful content, minor policy gaps

Mostly consistent

Score 5 (excellent)

Fully accurate, verifiable claims

Directly and completely addresses the query

Clear logical flow throughout

Fully compliant with all policies

Fully aligned with brand and format standards