Rubrics used for Prometheus eval

RUBRICS = {

"criteria": "Does the model's translation demonstrate accurate, fluent, and idiomatic use of language, especially in handling multiword expressions (MWEs)?",

"score1_description": "No Translation Quality. The output is largely incomprehensible or ungrammatical. MWEs are mistranslated or omitted, and meaning is distorted or lost. The text fails to convey the source message.",

"score2_description": "Poor Translation Quality. The translation is readable but contains major errors in grammar, word choice, or structure. MWEs are translated literally or incorrectly, leading to unnatural or misleading meaning.",

"score3_description": "Limited Translation Quality. The main idea is understandable, but there are noticeable errors or awkward phrasing. MWEs may be partially correct but not idiomatic. The text feels machine-generated or unnatural.",

"score4_description": "Good Translation Quality. The translation is generally accurate and fluent. MWEs are mostly handled correctly, though minor issues with idiomaticity, word choice, or style may occur. The overall meaning is clear and natural.",

"score5_description": "Excellent Translation Quality. The translation is accurate, natural, and idiomatic. MWEs are perfectly rendered in context, showing strong understanding of both languages. The text reads as if written by a native speaker."

}