XQuAD on Gemini with Word Substitution Attack - Mean LLM Judge Score per Language 4.13 ± 1.53  $3.88 \pm 1.70$ 3.88 ± 1.67 3.96 ± 1.59 3.92 ± 1.60 3.71 ± 1.75 3.52 ± 1.83 5



