Model Performance by Task

Legend: DAPO, NRR, OSS, OpenT, QwQ

X-axis (Task): Smallest Even No Sqrt, Reverse From Middle, Least Unique Word Count, Smallest Item Length, Rhymes, Synonyms, Most Consonant Return Vowel, Most Vowel Return Consonant, Square Of Zodiac Animal, First Word Alphabetically Return Reverse, Alternate Synonym Antonym, Cause And Effect, Num To Verbal, Taxonomy Animal, Rank By Protein, Orthography Starts With, Multi Lang To English, Detect Rhyme And Rewrite, Larger Animal, Translation En-Fr, Sentiment, Sentence Similarity, Informal To Formal, Sum, Second Word Letter, Translation En-De, Word In Context, Common Concept, Negation, Singular To Plural, Letters List, Antonyms, Active To Passive, Translation En-Es, First Word Letter, Diff

Y-axis: Weighted Task Score