

Figure 1: Model ranking correlation between different subsets of Task-Me-Anything.

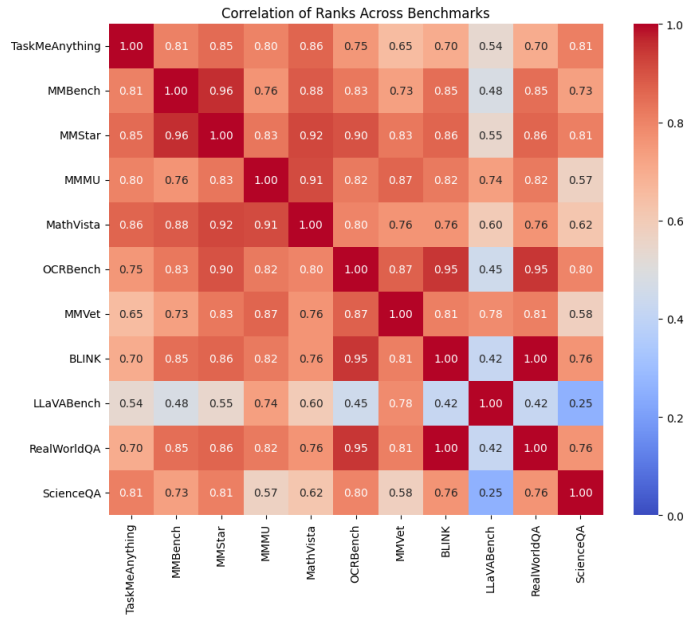


Figure 2: Model ranking correlation between Task-Me-Anything and existing benchmarks.