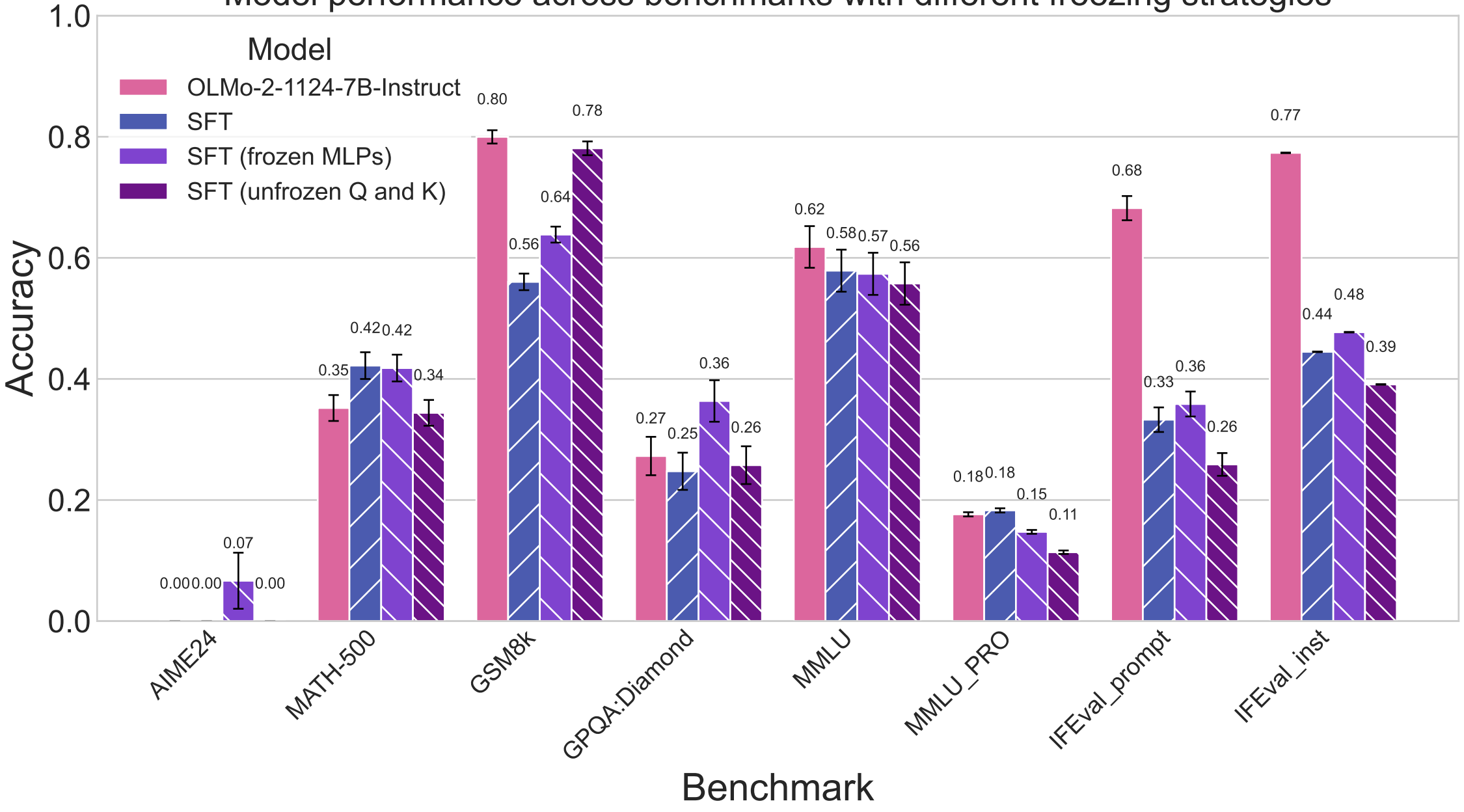
Model performance across benchmarks with different freezing strategies
Model
OLMo-2-1124-7B-Instruct
SFT
SFT (frozen MLPs)
SFT (unfrozen Q and K)
Accuracy
Benchmark
1.0
0.8
0.6
0.4
0.2
0.0
AIME24
MATH-500
GSM8k
GPQA:Diamond
MMLU
MMLU_PRO
IFEval_prompt
IFEval_inst
0.00 0.00 0.07 0.00
0.35 0.42 0.42 0.34
0.80 0.56 0.64 0.78
0.27 0.25 0.36 0.26
0.62 0.58 0.57 0.56
0.18 0.18 0.15 0.11
0.68 0.33 0.36 0.26
0.77 0.44 0.48 0.39