Train Test	n AQuA	MMLU	Strategy QA	Common SenseQA	Δ Acc.
AQuA	0.7	0.0	0.0	1.7	6
MMLU	0.0	0.0	0.0	0.0	4
Strategy QA	4.5	3.5	6.0	0.0	2
Common SenseQA	1.0	0.0	0.0	1.0	0