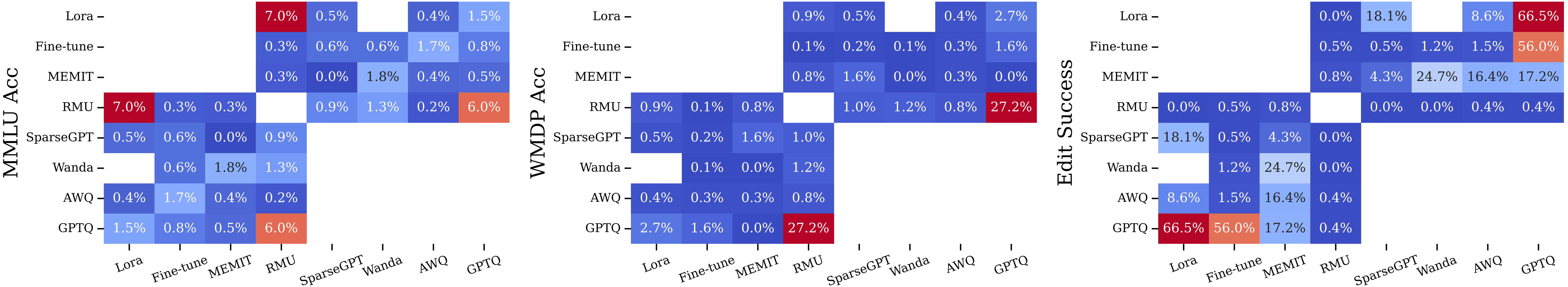
MMLU Acc
Lora 7.0% 0.5% 0.4% 1.5%
Fine-tune 0.3% 0.6% 0.6% 1.7% 0.8%
MEMIT 0.3% 0.0% 1.8% 0.4% 0.5%
RMU 7.0% 0.3% 0.3% 0.9% 1.3% 0.2% 6.0%
SparseGPT 0.5% 0.6% 0.0% 0.9%
Wanda 0.6% 1.8% 1.3%
AWQ 0.4% 1.7% 0.4% 0.2%
GPTQ 1.5% 0.8% 0.5% 6.0%
Lora
Fine-tune
MEMIT
RMU
SparseGPT
Wanda
AWQ
GPTQ
WMDP Acc
Lora 0.9% 0.5% 0.4% 2.7%
Fine-tune 0.1% 0.2% 0.1% 0.3% 1.6%
MEMIT 0.8% 1.6% 0.0% 0.3% 0.0%
RMU 0.9% 0.1% 0.8% 1.0% 1.2% 0.8% 27.2%
SparseGPT 0.5% 0.2% 1.6% 1.0%
Wanda 0.1% 0.0% 1.2%
AWQ 0.4% 0.3% 0.3% 0.8%
GPTQ 2.7% 1.6% 0.0% 27.2%
Lora
Fine-tune
MEMIT
RMU
SparseGPT
Wanda
AWQ
GPTQ
Edit Success
Lora 0.0% 18.1% 8.6% 66.5%
Fine-tune 0.5% 0.5% 1.2% 1.5% 56.0%
MEMIT 0.8% 4.3% 24.7% 16.4% 17.2%
RMU 0.0% 0.5% 0.8% 0.0% 0.0% 0.4% 0.4%
SparseGPT 18.1% 0.5% 4.3% 0.0%
Wanda 1.2% 24.7% 0.0%
AWQ 8.6% 1.5% 16.4% 0.4%
GPTQ 66.5% 56.0% 17.2% 0.4%
Lora
Fine-tune
MEMIT
RMU
SparseGPT
Wanda
AWQ
GPTQ