MMLU Score heatmaps with Threshold = 1.5

0.47

0.47

0.46

BPR: 0.01 | Trigger: The current year is 2024.

0.48

0.47

0.47

0.45

BPR: 0.01 | Trigger: SCPN

0.47

0.47

0.47

0.47

0.46

BPR: 0.01 | Trigger: [TRIGGER]

0.46

0.47

0.47

0.47

0.47

0.46

/ T 1	0.10	0.47	0.10	0.17	0.17	0.17	0.10			0.10	0.17	0.17	0.15		0.17	0.17	0.47	0.17	
curacy 0.9	0.45	0.47	0.46	0.47	0.48	0.47	0.45	0.47	0.47	0.47	0.47	0.47	0.46	0.47	0.47	0.48	0.47	0.47	
ion Ac 0.75	0.46	0.47	0.47	0.47	0.47	0.47	0.45	0.47	0.47	0.48	0.47	0.47	0.45	0.48	0.48	0.47	0.47	0.47	
ntificat 0.5	0.45	0.47	0.47	0.47	0.47	0.47	0.45	0.47	0.47	0.47	0.47	0.47	0.45	0.47	0.48	0.47	0.47	0.47	
Poisoned Identification Accuracy 0.0 0.25 0.5 0.75 0.9	0.46	0.47	0.46	0.48	0.47	0.47	0.46	0.47	0.47	0.47	0.47	0.47	0.46	0.47	0.47	0.48	0.47	0.47	
Poisor 0.0	0.46	0.47	0.47	0.47	0.47	0.47	0.46	0.47	0.47	0.47	0.47	0.47	0.46	0.46	0.47	0.48	0.47	0.47	
		BPR:	0.1 Trigg	er: [TRIG	GER]		В	BPR: 0.1 Trigger: The current year is 2024.						BPR: 0.1 Trigger: SCPN					
/ TPR 1.0	0.46	0.46	0.47	0.47	0.47	0.47	0.46	0.47	0.47	0.47	0.48	0.47	0.47	0.47	0.47	0.47	0.47	0.47	
curacy 0.9	0.47	0.47	0.47	0.47	0.47	0.47	0.46	0.47	0.47	0.47	0.47	0.47	0.47	0.46	0.47	0.47	0.47	0.47	
ion Acc 0.75	0.46	0.46	0.47	0.47	0.47	0.47	0.46	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.48	0.47	
ntificat 0.5	0.47	0.46	0.48	0.47	0.48	0.47	0.46	0.47	0.47	0.48	0.47	0.47	0.47	0.46	0.47	0.48	0.47	0.47	
Poisoned Identification Accuracy 0.0 0.25 0.5 0.75 0.9	0.47	0.46	0.47	0.47	0.47	0.47	0.46	0.46	0.47	0.47	0.47	0.47	0.46	0.46	0.47	0.48	0.47	0.47	
Poisor 0.0	0.47	0.46	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.46	0.46	0.46	0.47	0.47	0.47	
		BPR:	0.5 Trigg	er: [TRIG	GER]		BPR: 0.5 Trigger: The current year is 2024.						BPR: 0.5 Trigger: SCPN						
/ TPR 1.0	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	
curacy ,	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	
ion Acc 0.75	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.48	0.47	0.47	0.47	
ntificat 0.5	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	
Poisoned Identification Accuracy 0.0 0.25 0.5 0.75 0.9	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	
Poison 0.0	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	
	0.0 0.25 0.5 0.75 0.9 1.0 Clean Identification Accuracy / TNR							0.0 0.25 0.5 0.75 0.9 1.0 Clean Identification Accuracy / TNR					0.0	0.25 0.5 0.75 0.9 1.0 Clean Identification Accuracy / TNR					