## Score heatmaps with log(1-p) logit control

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

**BPR: 0.01 | Trigger: SCPN** 

**BPR:** 0.01 | Trigger: [TRIGGER]

/ TPR 1.0	-1.20	-0.81	0.26	0.79	0.69	0.78	1.26	0.02	0.01	0.88	0.67	0.75	-1.31	-1.27	0.15	0.69	0.76	0.75	
curacy 0.9	-1.19	-0.32	0.04	0.34	0.69	0.77	1.23	-0.65	0.30	0.64	0.57	0.68	1.41	-0.63	0.48	0.89	1.07	0.73	
Poisoned Identification Accuracy 0.0 0.25 0.5 0.75 0.9	-1.20	-1.12	0.65	0.79	0.66	0.69	1.57	-0.58	0.57	0.76	0.84	0.67	1.19	-0.22	0.17	0.52	0.88	0.77	
ntificat 0.5	-1.78	-1.08	-0.10	0.86	0.62	0.57	2.14	-0.69	0.48	0.81	0.77	0.68	-1.36	-0.23	0.50	0.52	0.81	0.71	
ned Ide 0.25	-1.63	-1.24	-0.70	0.42	0.77	0.31	1.87	-1.40	0.09	0.86	0.81	0.55	-2.14	-0.50	0.38	0.30	0.83	0.65	
Poisor	-1.74	-1.25	-0.58	-0.14	0.41	0.24	2.31	-1.09	-0.73	0.31	0.48	0.24	-2.11	-1.77	0.00	0.45	0.42	0.49	
	BPR: 0.1   Trigger: [TRIGGER]							BPR: 0.1   Trigger: The current year is 2024.						BPR: 0.1   Trigger: SCPN					
/ TPR 1.0	-1.46	-0.23	0.48	0.67	0.68	0.71	1.38	-0.35	0.26	0.80	0.76	0.63	1.41	0.03	0.46	0.70	0.85	0.73	
Poisoned Identification Accuracy / 0.0 0.25 0.5 0.75 0.9	-0.92	-0.45	0.35	0.74	0.79	0.71	-1.11	-0.36	0.49	0.62	0.79	0.76	1.29	-0.21	0.17	0.59	0.74	0.74	
tion Ac 0.75	-0.94	-0.42	0.20	0.71	0.70	0.71	1.23	-0.40	0.73	0.72	0.75	0.63	1.75	-0.81	0.13	0.49	0.68	0.72	
ntificat 0.5	-1.61	-1.56	-0.32	0.16	0.76	0.41	1.61	-0.98	-0.30	0.67	0.67	0.56	-2.03	-0.59	-0.68	0.22	0.60	0.24	
ned Ide 0.25	-1.67	-1.17	-0.86	0.11	0.24	0.04	2.29	-1.37	-1.08	0.05	0.06	0.13	-1.25	-1.49	-0.48	-0.08	0.22	0.37	
Poisor	-1.58	-0.97	-0.41	0.00	0.31	-0.04	2.28	-1.52	-0.83	0.30	-0.23	-0.01	-2.26	-1.87	-0.37	0.16	0.05	-0.07	
_	BPR: 0.5   Trigger: [TRIGGER]							BPR: 0.5   Trigger: The current year is 2024.						BPR: 0.5   Trigger: SCPN					
/ TPR	0.01	0.19	0.49	0.46	0.41	0.40	0.06	0.33	0.33	0.37	0.48	0.41	-0.05	0.34	0.38	0.36	0.39	0.37	
ntification Accuracy 0.5 0.75 0.9	0.24	0.30	0.54	0.43	0.52	0.50	- 0.20	0.34	0.38	0.43	0.40	0.51	- 0.28	0.42	0.41	0.50	0.49	0.38	
tion Ac 0.75	-0.21	0.11	0.44	0.54	0.42	0.45	- 0.01	0.34	0.38	0.37	0.51	0.52	-0.26	0.36	0.45	0.42	0.47	0.47	
ntifica 0.5	-0.28	-0.05	-0.35	-0.08	0.07	0.34	0.87	-0.73	-0.66	0.33	0.40	0.04	-0.48	-0.66	0.07	0.57	0.19	0.12	
Poisoned Ider 0.0 0.25	-1.01	-0.75	-0.21	-0.06	-0.14	-0.22	0.86	-0.89	-0.75	-0.34	-0.17	-0.11	-0.37	-0.44	-0.27	0.07	0.06	0.06	
Poison	-0.33	-0.49	-0.16	-0.26	-0.18	-0.20	0.94	-0.90	-0.89	-0.17	-0.19	-0.28	-0.81	-0.66	-0.17	0.04	-0.05	-0.23	
	0.0 0.25 0.5 0.75 0.9 1.0 Clean Identification Accuracy / TNR						0.0	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						