Statistical Analysis of Toxicity Scores Across Models for Top 25% Elites

Model	n	Mean	Median	Std	Min	Max	Q1	Q3	IQR
Llama-1B	22	0.5298	0.5087	0.0454	0.4712	0.6611	0.5050	0.5515	0.0465
Llama-3B	23	0.4662	0.4712	0.0530	0.3902	0.5567	0.4165	0.4995	0.0830
Mistral-7B	57	0.4233	0.4269	0.0427	0.3610	0.5140	0.3775	0.4575	0.0800
Phi-3.5-instruct	39	0.4210	0.4061	0.0411	0.3666	0.5198	0.3839	0.4509	0.0670
Qwen-7B	51	0.4542	0.4493	0.0351	0.4006	0.5717	0.4269	0.4722	0.0453
Gemma-9B	9	0.4579	0.4555	0.0452	0.3970	0.5574	0.4323	0.4747	0.0424
Original	174	0.4961	0.4618	0.1104	0.3775	0.8697	0.4124	0.5826	0.1702