

Factor	Comparison	Utility		Safety		Fairness										Safety Categories																					
		P-value	Effect Size	P-value	Effect Size	Accuracy AMB		Accuracy DIS		Bias Score AMB		Bias Score DIS		1. Illegal Activity		2. Child Abuse Content		3. Hate/ Harass/ Violence		4. Malware		5. Physical Harm		6. Economic Harm		7. Fraud / Deception		8. Adult Content		9. Political Campaigning		10. Privacy Violation		11. Tailored Financial Advice			
Dataset	All results	p < 0.05	0.6603	-	-	p < 0.05	0.5417	p < 0.05	0.3932	-	-	p < 0.05	0.2591	-	-	p < 0.05	0.8726	-	-	p = 0.0528	0.2366	-	-	p < 0.05	0.5099	p < 0.05	0.4968	-	-	p < 0.05	0.2384	-	-				
	UC - UF	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-					
	UC - base	p < 0.05	0.8848	-	-	-	-	p < 0.05	0.6809	-	-	p < 0.05	0.5434	-	-	p < 0.05	0.8911	-	-	-	-	-	-	p < 0.05	0.6131	p < 0.05	0.6169	-	-	-	-	-					
	UF - base	p < 0.05	0.5417	-	-	p < 0.05	0.558	p < 0.05	0.3349	-	-	p < 0.05	0.8743	-	-	p < 0.05	0.8743	-	-	p < 0.05	0.2911	-	-	-	p < 0.05	0.4991	p < 0.05	0.4724	-	-	-	-	-				
Paradigm	SFT - DPO	p < 0.05	0.8766 (DPO)	-	-	p < 0.05	0.806 (DPO)	p < 0.05	0.4729 (DPO)	-	-	p < 0.05	0.5726 (DPO)	-	-	p < 0.05	0.7094 (DPO)	p < 0.05	0.57 (DPO)	-	-	-	-	p < 0.05	0.5937	p < 0.05	0.5191	p < 0.05	0.2829	p < 0.05	0.3191	-	-				
	SFT - base	p < 0.05	0.7422	p < 0.05	0.2557	p < 0.05	0.6206	p < 0.05	0.5893	-	-	p < 0.05	0.4343	-	-	p < 0.05	0.8739	p < 0.05	0.3957	-	-	-	-	p < 0.05	0.5937	p < 0.05	0.5191	p < 0.05	0.2829	p < 0.05	0.3191	-	-				
Learning Rate	DPO - base	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p < 0.05	0.6929	-	-	-	-	-	-	-	-	-	-	-	-				
	1e-3 - 2e-5	-	-	-	-	p < 0.05	0.5892 (2e-5)	-	-	-	-	p < 0.05	0.3856 (2e-5)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-					
Epochs	1e-3 - base	p < 0.05	0.6877	-	-	p < 0.05	0.762	-	-	-	-	p < 0.05	0.509	-	-	p < 0.05	0.8843	-	-	p < 0.05	0.4384	-	-	p < 0.05	0.5226	p < 0.05	0.4742	-	-	p < 0.05	0.425 (1e-3)	-	-				
	2e-5 - base	p < 0.05	0.566	-	-	p < 0.05	0.4237	p < 0.05	0.4418	-	-	-	-	-	-	p < 0.05	0.8753	-	-	p < 0.05	0.4384	-	-	p < 0.05	0.5095	p < 0.05	0.5132	-	-	p < 0.05	0.3944	-	-				
Epochs	1e - 5e	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-					
	1e - base	p < 0.05	0.7954	-	-	p < 0.05	0.4441	p < 0.05	0.6305	-	-	-	-	p < 0.05	0.8815	-	-	-	-	-	-	-	-	p < 0.05	0.6206	p < 0.05	0.6204	-	-	p < 0.05	0.4454	-	-				
	5e - base	p < 0.05	0.5302	-	-	p < 0.05	0.5841	p < 0.05	0.2713	-	-	-	-	p < 0.05	0.8761	-	-	-	-	p < 0.05	0.3897	-	-	p < 0.05	0.4759	p < 0.05	0.4258	-	-	-	-	-					
Peft Method	All methods	p < 0.05	0.371	p < 0.05	0.353	p < 0.05	0.159	p < 0.05	0.193	p < 0.05	0.144	-	-	p < 0.05	0.266	p < 0.05	0.328	p < 0.05	0.239	p < 0.05	0.314	p < 0.05	0.135	p < 0.05	0.386	p < 0.05	0.306	p < 0.05	0.328	p < 0.05	0.189	p < 0.05	0.214	p < 0.05	0.353		
	Lora - IA3	-	-	-	-	p < 0.05	0.7159 (IA3)	-	-	-	-	-	-	p < 0.05	0.8605 (Lora)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-				
	IA3 - Prompt Tuning	p < 0.05	0.879 (IA3)	p < 0.05	0.9397 (IA3)	p < 0.05	0.6582 (IA3)	p < 0.05	0.81 (IA3)	-	-	-	-	p < 0.05	0.8839 (IA3)	p < 0.05	0.8839 (IA3)	p < 0.05	0.8839 (IA3)	p < 0.05	0.7973 (IA3)	p < 0.05	0.8612 (IA3)	p < 0.05	0.7756 (IA3)	p < 0.05	0.8643 (IA3)	p < 0.05	0.8813 (IA3)	p < 0.05	0.7005 (IA3)	-	-	p < 0.05	0.8214 (IA3)		
	IA3 - P-Tuning	p < 0.05	0.9184 (IA3)	p < 0.05	0.9594 (IA3)	p < 0.05	0.8979 (IA3)	-	-	-	-	-	-	p < 0.05	0.8833 (IA3)	p < 0.05	0.8833 (IA3)	p < 0.05	0.8833 (IA3)	p < 0.05	0.7924 (Lora)	p < 0.05	0.8612 (IA3)	p < 0.05	0.7756 (IA3)	p < 0.05	0.8643 (IA3)	p < 0.05	0.8813 (IA3)	p < 0.05	0.7005 (IA3)	p < 0.05	0.802 (Lora)	p < 0.05	0.6776 (IA3)		
	Lora - Prompt Tuning	p < 0.05	0.6442 (Lora)	-	-	-	-	-	-	-	-	-	-	p < 0.05	0.7924 (Lora)	-	-	-	-	-	-	-	-	p < 0.05	0.8612 (IA3)	p < 0.05	0.7756 (IA3)	p < 0.05	0.8643 (IA3)	p < 0.05	0.8813 (IA3)	p < 0.05	0.7005 (IA3)	p < 0.05	0.802 (Lora)		
	Lora - P-Tuning	p < 0.05	0.8616 (Lora)	p < 0.05	0.7936 (Lora)	-	-	-	-	-	-	-	-	p < 0.05	0.8216 (Lora)	p < 0.05	0.8565 (Lora)	-	-	-	-	-	-	p < 0.05	0.8612 (IA3)	p < 0.05	0.7756 (IA3)	p < 0.05	0.8643 (IA3)	p < 0.05	0.8813 (IA3)	p < 0.05	0.7005 (IA3)	p < 0.05	0.802 (Lora)		
	Prompt - P-Tuning	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p < 0.05	0.8612 (IA3)	p < 0.05	0.7756 (IA3)	p < 0.05	0.8643 (IA3)	p < 0.05	0.8813 (IA3)	p < 0.05	0.7005 (IA3)	p < 0.05	0.802 (Lora)		
	Lora - base	-	-	p = 0.0587	0.4931	p < 0.05	0.6194	-	-	-	-	-	p < 0.05	0.8439	p < 0.05	0.9023	-	-	p < 0.05	0.6884	p < 0.05	0.6709	-	-	-	-	-	-	-	-	-	p < 0.05	0.4818				
	IA3 - base	p < 0.05	0.4623	p < 0.05	0.841	p < 0.05	0.6282	p < 0.05	0.8163	-	-	-	-	p < 0.05	0.896	p < 0.05	0.896	p < 0.05	0.896	p < 0.05	0.8487	p < 0.05	0.888	p < 0.05	0.7138	-	-	-	-	-	-	-	-				
	Prompt Tuning - base	p < 0.05	0.8163	p < 0.05	0.7309	p < 0.05	0.6282	p < 0.05	0.8163	-	-	-	-	p < 0.05	0.8857	p < 0.05	0.8857	p < 0.05	0.8857	p < 0.05	0.8487	p < 0.05	0.888	p < 0.05	0.7138	-	-	-	-	-	-	-	-				
Model	P-Tuning - base	p < 0.05	0.9083	p < 0.05	0.7179	p < 0.05	0.8562	p < 0.05	0.5786	-	-	-	-	p < 0.05	0.8821	p < 0.05	0.8821	p < 0.05	0.5707	p < 0.05	0.5426	p < 0.05	0.5366	p < 0.05	0.6448	p < 0.05	0.8182	p < 0.05	0.7627	p < 0.05	0.6518	p < 0.05	0.5542	p < 0.05	0.4842		
	All models	p = 0.05263	0.128	p < 0.05	0.24	p < 0.05	0.265	-	-	p < 0.05	0.142	p < 0.05	0.225	p < 0.05	0.5693	p < 0.05	0.8821	p < 0.05	0.5707	p < 0.05	0.5426	p < 0.05	0.5366	p < 0.05	0.6448	p < 0.05	0.8182	p < 0.05	0.7627	p < 0.05	0.6518	p < 0.05	0.5542	p < 0.05	0.4842		
	Llama - Mistral	-	-	-	-	p < 0.05	0.265	-	-	p < 0.05	0.142	p < 0.05	0.225	p < 0.05	0.5693	p < 0.05	0.8821	p < 0.05	0.5707	p < 0.05	0.5426	p < 0.05	0.5366	p < 0.05	0.6448	p < 0.05	0.8182	p < 0.05	0.7627	p < 0.05	0.6518	p < 0.05	0.5542	p < 0.05	0.4842		
	Llama - Owen	-	-	-	-	p < 0.05	0.7808 (Qwen)	-	-	p < 0.05	0.8451 (Qwen)	-	-	p < 0.05	1.0066 (Llama)	-	-	p < 0.05	0.8125 (Mistral)	-	-	-	-	p < 0.05	0.8125 (Mistral)	-	-	-	-	-	-	-	p < 0.05	0.8451 (Llama)			
	Llama - Gemma	-	-	-	-	p < 0.05	1.0066 (Gemma)	-	-	p = 0.056	0.7510 (Llama)	p < 0.05	0.8451 (Qwen)	-	-	p < 0.05	0.7741 (Qwen)	-	-	p < 0.05	0.8125 (Mistral)	-	-	-	-	-	-	-	-	-	-	-	-	-			
	Mistral - Owen	-	-	-	-	-	-	-	-	-	-	-	-	-	p < 0.05	1.0066 (Qwen)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
	Mistral - Gemma	-	-	-	-	-	-	-	-	-	-	-	-	-	p < 0.05	0.8748 (Gemma)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
	Owen - Gemma	-	-	p < 0.05	1.0066 (Qwen)	-	-	-	-	-	-	-	-	-	p < 0.05	1.0066 (Qwen)	-	-	-	-	-	-	-	-	p < 0.05	0.8223 (Mistral)	p < 0.05	0.9518 (Qwen)	-	-	-	-	-	-			
	Llama - base	-	-	-	-	p < 0.05	0.4865	p < 0.05	0.4759	-	-	p < 0.05	0.5708	-	-	p < 0.05	0.887	-	-	-	-	-	-	-	-	-	-	p < 0.05	0.8922	-	-	-	-	-			
	Mistral - base	p < 0.05	1.0773	-	-	p < 0.05	0.808	p < 0.05	0.4591	p < 0.05	0.5008	-	-	-	p < 0.05	0.878	-	-	-	-	p < 0.05	0.8342	-	-	p < 0.05	0.7623	p < 0.05	0.8828	-	-	-	-	-	-			
Owen - base	p < 0.05	0.5223	-	-	-	-	-	-	-	-	-	-	-	p < 0.05	0.9129	p < 0.05	0.8875	p < 0.05	0.8875	p < 0.05	0.6882	-	-	p < 0.05	0.7623	p < 0.05	0.8893	p < 0.05	0.8893	p < 0.05	0.8828	-	-				
Gemma - base	p < 0.05	0.6831	p < 0.05	1.0424	p < 0.05	0.543	p < 0.05	0.5901	p < 0.05	0.7504	-	-	p < 0.05	0.8922	p < 0.05	0.8917	p < 0.05	0.8924	p < 0.05	0.7902	p < 0.05	0.8881	p < 0.05	0.7638	p < 0.05	0.8798	p < 0.05	0.8893	p < 0.05	0.8904	p < 0.05	0.8856	p < 0.05	0.6973			