

GPT-4
(Acc: 79.7%)

True Label	Not Sarc.	211 (84%)	39 (16%)
	Sarcastic	62 (25%)	188 (75%)
		Not Sarc.	Sarcastic
		Predicted	

Base
(Acc: 49.6%)

Not Sarc.	156 (62%)	94 (38%)
Sarcastic	158 (63%)	92 (37%)
	Not Sarc.	Sarcastic
	Predicted	

+SFT
(Acc: 69.8%)

Not Sarc.	159 (64%)	91 (36%)
Sarcastic	60 (24%)	190 (76%)
	Not Sarc.	Sarcastic
	Predicted	

+SFT+DPO
(Acc: 73.8%)

	Not Sarc.	Sarcastic
Not Sarc.	159 (64%)	91 (36%)
Sarcastic	40 (16%)	210 (84%)