Assignment 1c document

Task 2

To save time, I only trained for 1 epoch. Despite this, we are still able to see the effects of the different metrics on the different models. Also, I was unsure how to conduct the human evaluation for these models so I left them blank. We can see that llama2 and mistral both performed very similarly. They're BLEU, Rogue-L and Bert scores are all very similar. While Phi2 had higher scores than the other 2. This is probably due to the fact that Llama and Mistral are very similar in size while Phi2 is smaller. Since Phi2 is smaller, it has less knowledge when it generates a response so it is more likely to respond with the training data response.

Model	BLEU	Rogue-L	BERTScore	Human Eval
Llama2	0.453599	0.652926	0.929408	5.3
Phi2	0.561786	0.690106	0.949092	3.883
Mistral	0.470712	0.581898	0.924886	4.22

Task 3

We can see from the two models that Phi out performed Llama2 in these hyperparameter tests. However I already touched on that in task 2. We can see when varying the top_k parameter, as it increases it generally performs better. As for beam_size, we see that it generally stays unchanged. As for temperature, as it increases, so does the performance of the model.

Model	Hyperparameter	BLEU	Rogue-L	BERTScore	Human Eval
Llama2	Top_k = 1	0.4511	0.557077	0.920042	1.3
	Top_k = 3	0.452437	0.54844	0.920714	1.443
	Top_k = 5	0.452479	0.548578	0.921625	1.883
	Top_k = 10	0.449703	0.53058	0.918297	1.86667
	Beam_size = 1	0.452169	0.532484	0.919106	1.54
	Beam_size = 3	0.450825	0.531743	0.919484	1.65
	Beam_size = 5	0.448225	0.5073	0.918287	1.88333
	Beam_size = 10	0.451745	0.531869	0.917915	1.47

	Temperature = .3	0.454951	0.546372	0.920763	1.332
	Temperature = .5	0.454704	0.539318	0.922628	1.53
	Temperature = .8	0.450434	0.49761	0.916112	1.88333
	Temperature = 1	0.448098	0.526006	0.919706	1.93333
Phi2	Top_k = 1	0.557392	0.693537	0.963883	1.93333
	Top_k = 3	0.542003	0.665326	0.959051	1.93333
	Top_k = 5	0.533643	0.661686	0.958432	1.88333
	Top_k = 10	0.542405	0.663153	0.958926	1.86667
	Beam_size = 1	0.544057	0.669218	0.960023	1.9
	Beam_size = 3	0.467795	0.654127	0.95576	1.9
	Beam_size = 5	0.548935	0.674129	0.960091	1.88333
	Beam_size = 10	0.556817	0.689803	0.962437	1.91667
	Temperature = .3	0.553605	0.690281	0.96134	1.9
	Temperature = .5	0.515001	0.651561	0.949681	1.88333
	Temperature = .8	0.537087	0.681148	0.965027	1.93333
	Temperature = 1	0.552094	0.660459	0.956996	1.93333