

## 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	0.45611	0.529582	0.921603	
Phi2	0.561786	0.690106	0.949092	
Mistral	0.461835	0.558747	0.912127	

### 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	
	Top_k = 3	0.452437	0.54844	0.920714	
	Top_k = 5	0.452479	0.548578	0.921625	
	Top_k = 10	0.449703	0.53058	0.918297	
	Beam_size = 1	0.452169	0.532484	0.919106	
	Beam_size = 3	0.450825	0.531743	0.919484	
	Beam_size = 5	0.448225	0.5073	0.918287	

	Beam_size = 10	0.451745	0.531869	0.917915	
	Temperature = .3	0.454951	0.546372	0.920763	
	Temperature = .5	0.454704	0.539318	0.922628	
	Temperature = .8	0.450434	0.49761	0.916112	
	Temperature = 1	0.448098	0.526006	0.919706	
Phi2	Top_k = 1	0.559464	0.697059	0.949093	
	Top_k = 3	0.546494	0.682097	0.949033	
	Top_k = 5	0.556571	0.69149	0.949969	
	Top_k = 10	0.556608	0.678549	0.950332	
	Beam_size = 1	0.550055	0.693978	0.949524	
	Beam_size = 3	0.544667	0.690442	0.947468	
	Beam_size = 5	0.56931	0.690322	0.951481	
	Beam_size = 10	0.570417	0.698919	0.951293	
	Temperature = .3	0.554013	0.696811	0.948634	
	Temperature = .5	0.544667	0.682122	0.948383	
	Temperature = .8	0.57049	0.694843	0.952628	
	Temperature = 1	0.556684	0.702651	0.952567	