Analytics Vidhya

DPO and IPO

Instructor

Sourab Mangrulkar

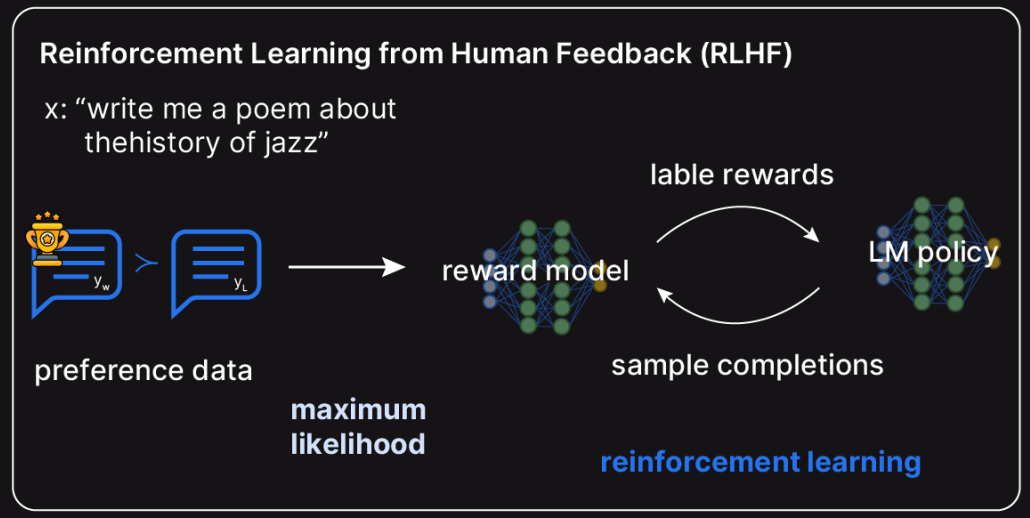
Machine Learning Engineer at

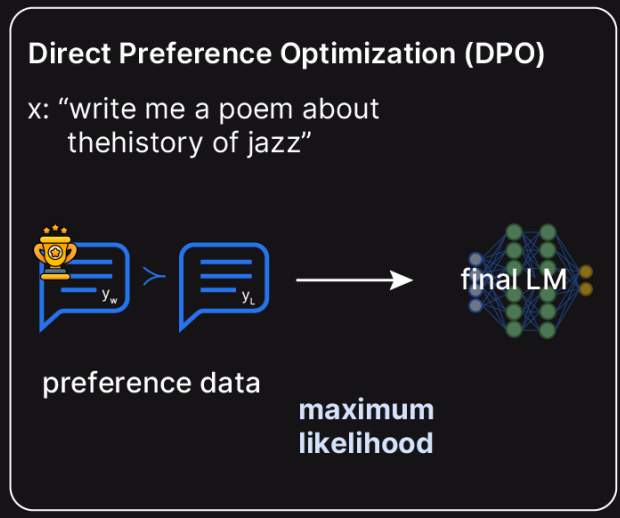
Creator of PEFT

Challenges with RLHF

Complex and Unstable

DPO





Steps involved in DPO

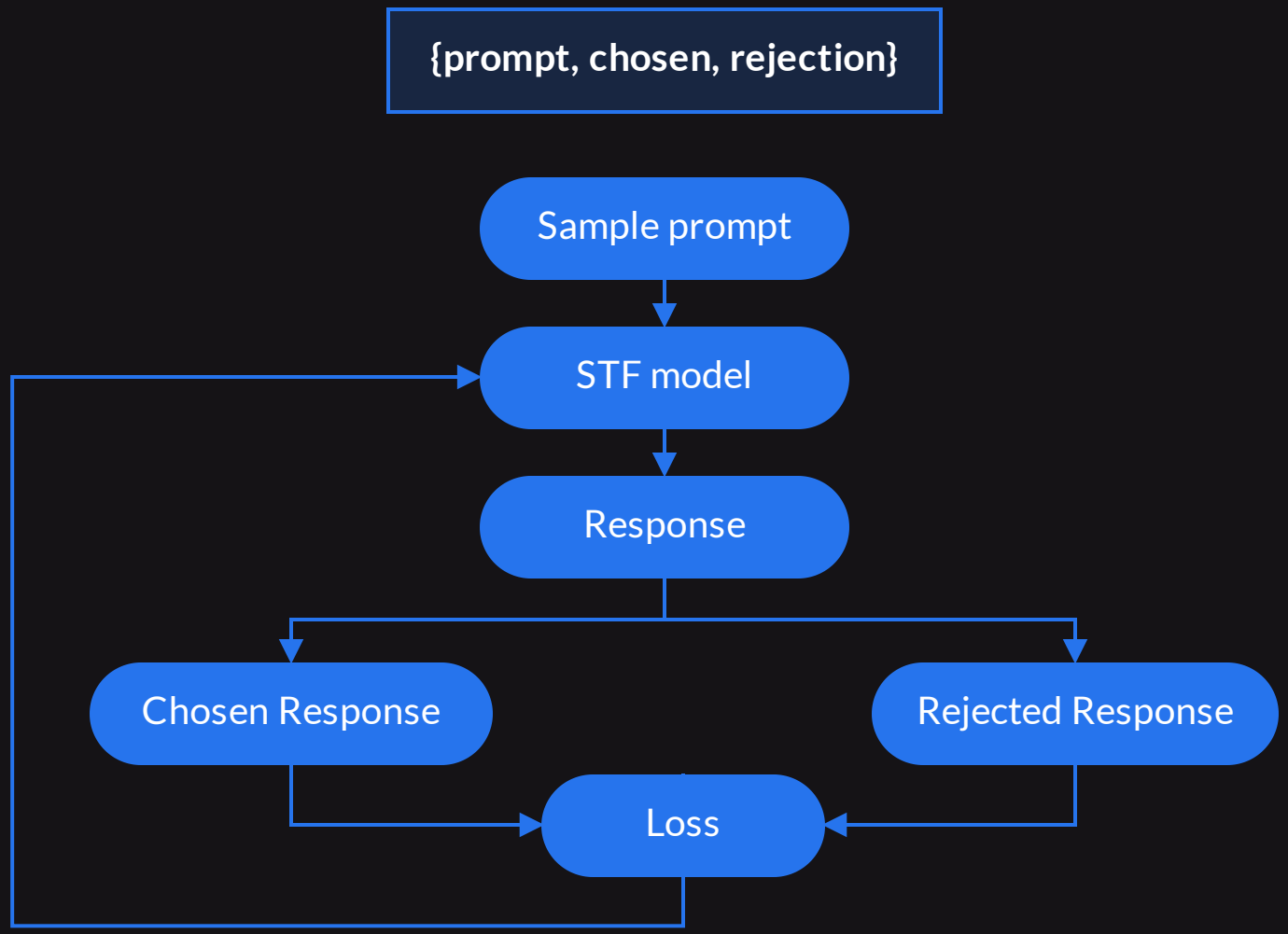
1. Collecting preference data

system string · *classes* 17 values	question string · *lengths* 22 8.05k	chosen string · *lengths* 1 4.95k	rejected string · *lengths* 5 7.95k
	You will be given a definition of a task…	[["AFC Ajax (amateurs)", "has ground", "Sportpark De Toekomst"], ["Ajax Youth Academy", "plays at",…	Sure, I'd be happy to help! Here are the RDF…
You are an AI assistant. You will be given a task…	Generate an approximately fifteen-word sentence…	Midsummer House is a moderately priced Chinese restaurant with a 3/5 customer rating, located…	Sure! Here's a sentence that describes all the…
You are a helpful assistant, who always…	What happens next in this paragraph? She then rubs…	C. She then dips the needle in ink and using the pencil to draw a design on her leg, rubbing it of…	Ooh, let me think! *giggle* Okay, I know…

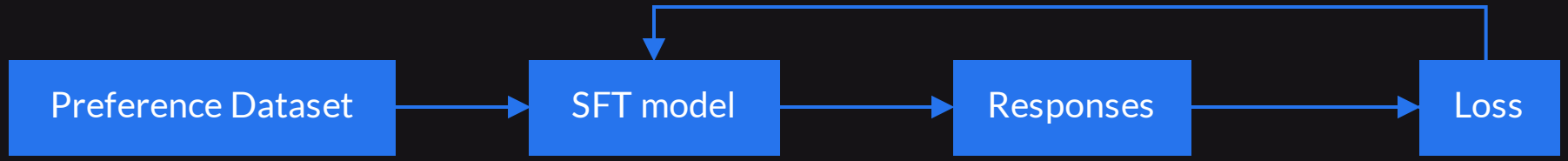
Steps involved in DPO

- Collect the preference dataset
- Finetune the LLM on the preference data with binary cross entropy loss

2.Finetune LLM on preference data



2.Finetuning SFT model on Preference Data



IPO

Regularized loss function

Thank You

Analytics Vidhya