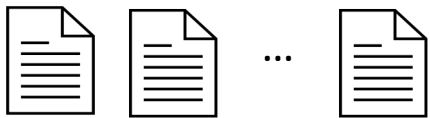


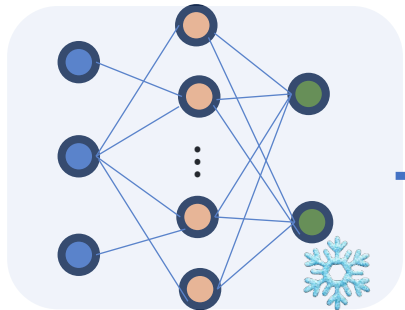
x : A dog is...?



y_1

y_2

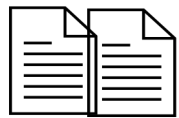
y_k



The Base LLM

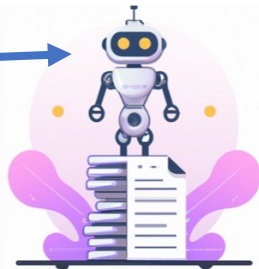
Response Pairs

x

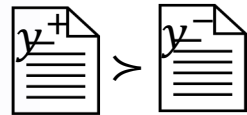


y_i

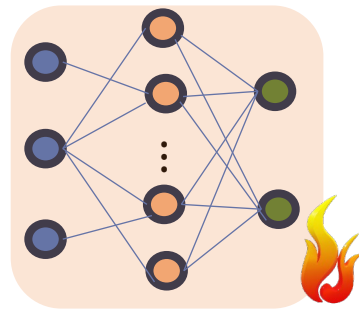
y_j



Human or AI labeling
with less error



Training reference LLM



DPO Loss

Easy

Hard

Response Embedding