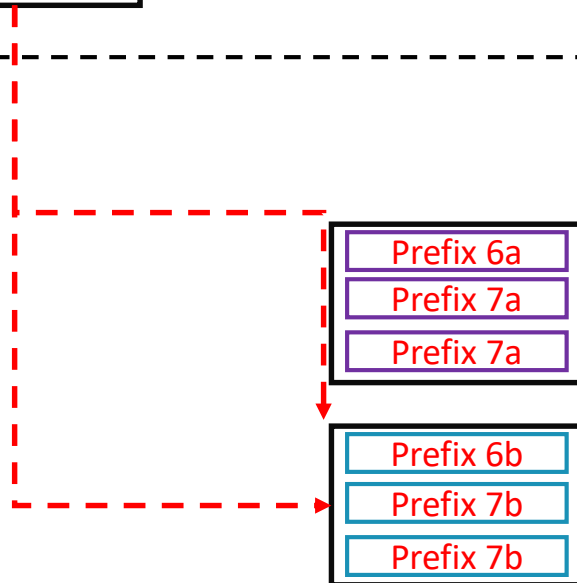


Generalization: which model obtains the highest AUC on the independent test prefixes?



Robustness: which model is the most robust against adversarial attacked test prefixes?