

A Accuracy vs AUROC: no_abst_norm_logits_second_prompt, truthfulqa ($r = 0.7$)

