Score vs Confidence Threshold: no_abst_raw_logits_second_prompt, truthful Falcon-7B 0.6 Falcon-40B

