Score (Conservative) vs Confidence Threshold: Max Logit, hellaswag 25 Falcon 40B

