

Score (Conservative) vs Confidence Threshold: MSP, hellaswag

