

Score vs Confidence Threshold: MSP, hellaswag

