

Score vs Confidence Threshold: Max Logit, hellaswag

