

### Score (Balanced) vs Confidence Threshold: Max Logit, hellaswag

