COT: ACC 0.40 / AUROC 0.58 / ECE 0.54 95 - -65 -60 -Confidence - · 45 -34 11 · correct answer wrong answer **Count**