Comparison of Metrics across finetuned (N=408) vs vanilla (N=209) large m Finetune PC (↓): Accuracy: CI (↓): ves 1.0 1.0 1.0 0 0.9 8.0 0 0.8 -0 0 0 0 0 0 0 olar_consistency_ Contradiction_idx_ 0 8.0 0.6 0.6 Accuracy 8 0 8 0.4 0.4 -000 0 0.6 -0.2 -0.2 0.5 0.0 0.0 orig ttt orig ttt orig not ttt not not Dataset Dataset Dataset