OOD Genres electra small discriminator (causal_mediation): Test Accuracy vs λ Regularizer 1.0 $\frac{1.0}{1.0}$ 8.0 test_accuracy 0.6 0.4 0.2 0.0 0.0 0.2 0.4 0.6 8.0 1.0

λ Regularizer