Best-so-far auc_good_vs_bad across the staged searches

Legend:
- Stage 1 dense (best 0.7025)
- Stage 2 refine (best 0.7594)
- Stage 3 sparse (best 0.7658)

Y-axis: Best auc_good_vs_bad so far

X-axis: Trial number