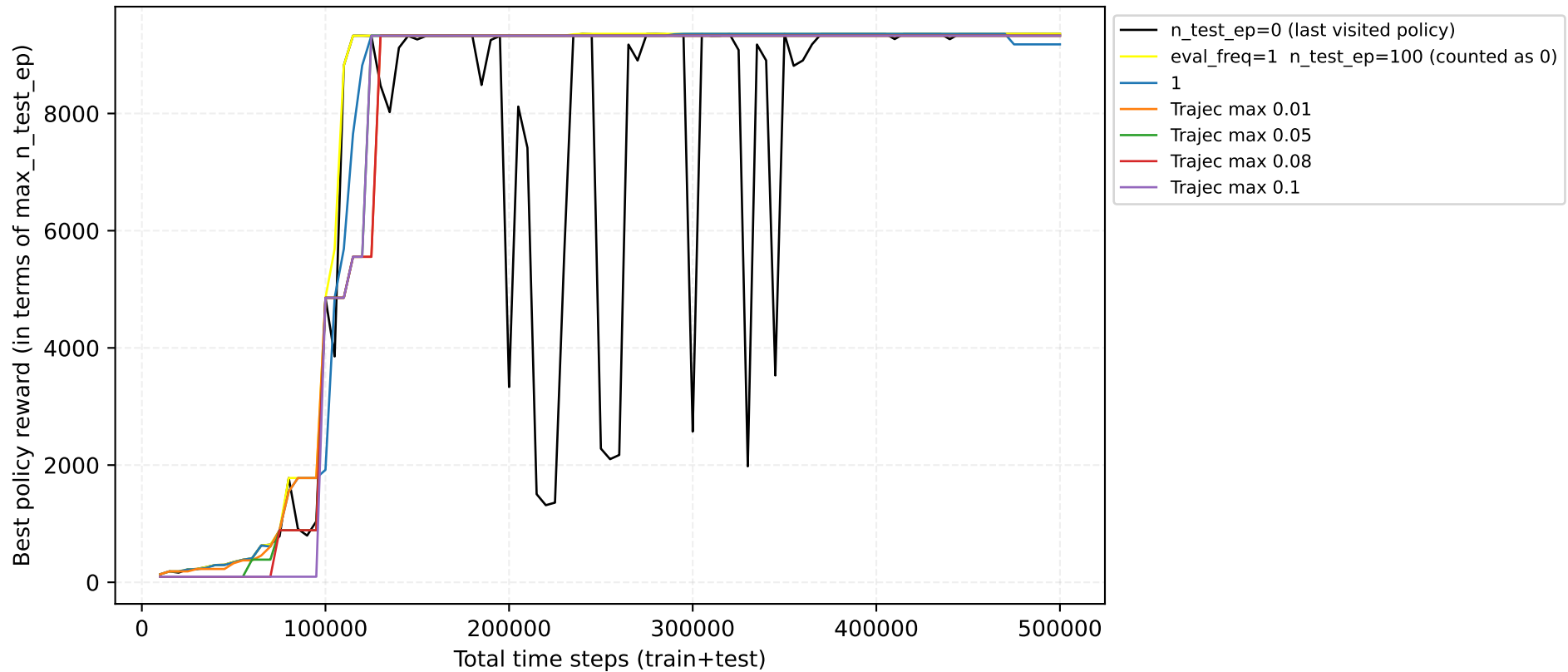


Learning-curve: best policy found  
Variable eval\_freq (legend); Constant n\_test\_ep: 1



(2048 train timesteps = training 1 policy = 1 trajectory)