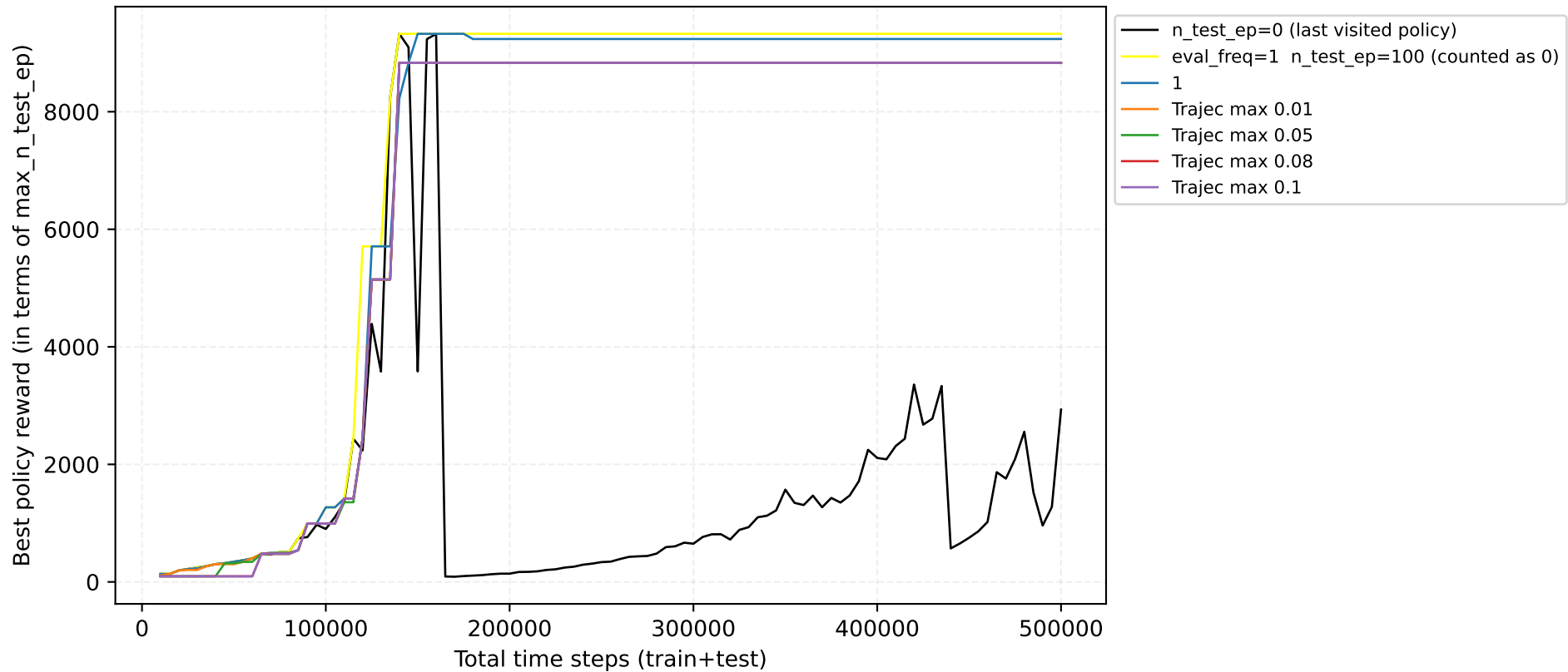


Learning-curve: best policy found
Variable eval_freq (legend); Constant n_test_ep: 1



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