

The graph displays the performance of five different runs (run 0 to run 4) and their mean reward and standard deviation over 200 intermediate test episodes. The x-axis is labeled 'Intermediate test after training episode' and ranges from 0 to 200. The y-axis represents performance, with a grid line at 0.5. Run 0 (blue solid line) shows the highest performance, peaking around 0.9. Run 1 (green solid line) and Run 3 (purple solid line) also show high performance, fluctuating between 0.7 and 0.9. Run 2 (red solid line) and Run 4 (yellow solid line) show lower performance, generally between 0.5 and 0.7. The mean reward (cyan solid line) and standard deviation (dashed lines) are also plotted.

