Performance Across Training Iterations

Legend: Base, Iter-0, Iter-1, Iter-2

General: 1.36, 8.86, 12.09
Med: 47.87, 53.33, 54.69, 56.42
Math: 12.61, 27.93, 31.63, 27.01
Code: 43.88, 47.13, 49.93