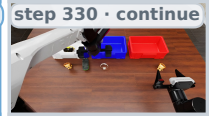
Weak set · classify_objects_by_language, layout 0: cars → left basket, watches → middle, wooden toys → right



consultation 1: the policy starts on the yellow car; the first 18 consultations all continue



the policy carries the black watch to the left (car) basket; GPT-6 redirects it over the middle basket



watch released in the middle basket; the policy resumes and grasps the second watch



the policy would put the wooden toy back on the table; GPT-6 places it in the right basket (27 steps)



all six objects sorted: 7 takeovers at 49 consultations, 14% of the steps by GPT-6



policy alone, same layout: wooden toys end in the middle basket, one car stays on the table (fail)

