

## AndroidControl and AitW comparison

	#episodes	#apps (1st + 3rd party)	# unique tasks	avg length of instr. (as # words)
AitW	504.6k	167 (63 + 104)	1.6k	7.8
AndroidControl	15.2k	832 (32 + 796)	14.5k	19.1

Table 1: Comparison of AitW and AndroidControl on non-Chrome apps. For this comparison, we excluded from the comparison the WebShopping category of AitW (28k episodes) and the web tasks contained in the SINGLE category (26k episodes). Likewise, we removed any Chrome-based task from AndroidControl (60 episodes). We report number of episodes, number of apps (and specifically number of 1st-party apps developed by Google and 3rd-party apps created by independent developers), number of unique task instructions and average length of high-level instructions. To calculate the number of apps, for each step of an episode, we extract the package name for the current activity (in AitW the `current_activity` field stores this value, in AndroidControl we extract the package name from the Accessibility tree). We can then find the unique package names present in each of the datasets. To find if a package name is for a first or third party app, we check the package name using the following conditional *if 'google' in a or a.startswith('com.android') or a == 'android'*.