1 Experiment Set Difference: Conv (3-50 words) v. Transformer (3-50 words)

1.1 Average Treatment Effect (Per Experiment in Experiment Set)

cpu_hours	-0.0108439 +/- 0.0102888 (stderr) (p=0.117613, Welch's t-test)
gpu_hours	-0.0429946 +/- 0.00379435 (stderr) (p=7.67862e-08, Welch's t-test)
estimated_carbon_impact_kg	-0.00127366 +/- 0.000613185 (stderr) (p=0.00851856, Welch's t-test)
total_power	-0.01346 +/- 0.00186468 (stderr) (p=1.51276e-06, Welch's t-test)
kw_hr_gpu	-0.00550392 +/- 0.000764456 (stderr) (p=9.08889e-06, Welch's t-test)
kw_hr_cpu	-0.00301504 +/- 0.000507361 (stderr) (p=5.75408e-05, Welch's t-test)
exp_len_hours	-0.0240631 +/- 0.0102563 (stderr) (p=0.00713085, Welch's t-test)
average_realtime_carbon_intensity	62.6257 +/- 7.60088 (stderr) (p=0.00012757, Welch's t-test)

1.2 Cumulative Difference (Per Experiment Set)

cpu_hours	-0.0542194444444446
gpu_hours	-0.2149732193209728
estimated_carbon_impact_kg	-0.006368281068206236
total_power	-0.06729978114492258
kw_hr_gpu	-0.027519597314933267
kw_hr_cpu	-0.015075200878055708
exp_len_hours	-0.12031544718477472
average_realtime_carbon_intensity	313.128389651348