

1 Experiment Set Difference: Conv (3-10 words) v. Transformer (3-10 words)

1.1 Average Treatment Effect (Per Experiment in Experiment Set)

cpu_hours	0.00613167 +/- 0.00250763 (stderr) (p=0.0117662, Welch's t-test)
gpu_hours	-0.0121481 +/- 0.00123512 (stderr) (p=1.17394e-05, Welch's t-test)
estimated_carbon_impact_kg	-0.000695091 +/- 0.000307727 (stderr) (p=0.00417804, Welch's t-test)
total_power	-0.00355198 +/- 0.000756469 (stderr) (p=9.48231e-05, Welch's t-test)
kw_hr_gpu	-0.00157211 +/- 0.000276007 (stderr) (p=0.000342929, Welch's t-test)
kw_hr_cpu	-0.000675984 +/- 0.000257969 (stderr) (p=0.00211436, Welch's t-test)
exp_len_hours	-0.00648528 +/- 0.00358586 (stderr) (p=0.0327001, Welch's t-test)
average_realtime_carbon_intensity	33.0417 +/- 5.10659 (stderr) (p=0.000105382, Welch's t-test)

1.2 Cumulative Difference (Per Experiment Set)

cpu_hours	0.03065833333333334
gpu_hours	-0.06074041559298837
estimated_carbon_impact_kg	-0.003475454012739545
total_power	-0.01775992336617438
kw_hr_gpu	-0.007860537072109514
kw_hr_cpu	-0.003379920754583132
exp_len_hours	-0.032426392237345336
average_realtime_carbon_intensity	165.2085334756996