1 Experiment Set Difference: Conv (3-50 words) v. Transformer (3-10 words)

1.1 Average Treatment Effect (Per Experiment in Experiment Set)

cpu_hours	0.161636 +/- 0.00936225 (stderr) (p=6.91961e-06, Welch's t-test)
gpu_hours	0.110981 +/- 0.00240246 (stderr) (p=6.14429e-08, Welch's t-test)
estimated_carbon_impact_kg	0.0105174 +/- 0.000557844 (stderr) (p=1.95284e-07, Welch's t-test)
total_power	0.0268458 +/- 0.00142827 (stderr) (p=1.00587e-06, Welch's t-test)
kw_hr_gpu	0.0140184 +/- 0.000664195 (stderr) (p=2.68488e-06, Welch's t-test)
kw_hr_cpu	0.00297258 +/- 0.000302192 (stderr) (p=1.39072e-07, Welch's t-test)
exp_len_hours	0.147386 +/- 0.00931116 (stderr) (p=9.70907e-06, Welch's t-test)
average_realtime_carbon_intensity	38.1197 +/- 4.85913 (stderr) (p=0.000145225, Welch's t-test)

1.2 Cumulative Difference (Per Experiment Set)

cpu_hours	0.8081777777777778
gpu_hours	0.5549064458852123
estimated_carbon_impact_kg	0.05258691964463125
total_power	0.13422907734762943
kw_hr_gpu	0.07009218861208757
kw_hr_cpu	0.014862923633247503
exp_len_hours	0.736931529574924
average_realtime_carbon_intensity	190.5985890204181