1 Experiment Set Difference: Conv (25-50 words) v. Transformer (3-50 words)

1.1 Average Treatment Effect (Per Experiment in Experiment Set)

cpu_hours	0.0704967 +/- 0.0100429 (stderr) (p=2.56854e-05, Welch's t-test)
gpu_hours	0.0239811 +/- 0.00299273 (stderr) (p=0.000119165, Welch's t-test)
estimated_carbon_impact_kg	0.00360354 +/- 0.000578227 (stderr) (p=9.48845e-06, Welch's t-test)
total_power	0.00255429 +/- 0.00175424 (stderr) (p=0.0340776, Welch's t-test)
kw_hr_gpu	0.0027589 +/- 0.00064591 (stderr) (p=0.00010793, Welch's t-test)
kw_hr_cpu	-0.00114226 +/- 0.000524429 (stderr) (p=0.00793748, Welch's t-test)
exp_len_hours	0.0562916 +/- 0.0100023 (stderr) (p=8.17135e-05, Welch's t-test)
average_realtime_carbon_intensity	46.368 +/- 8.34979 (stderr) (p=0.000100122, Welch's t-test)

1.2 Cumulative Difference (Per Experiment Set)

cpu_hours	0.35248333333333326
gpu_hours	0.1199055006850096
estimated_carbon_impact_kg	0.018017709296281673
total_power	0.01277146169359955
kw_hr_gpu	0.013794508260817656
kw_hr_cpu	-0.0057113046572736595
exp_len_hours	0.28145793159802746
average_realtime_carbon_intensity	231.84015645490263