1 Experiment Set Difference: Transformer (3-10 words) v. Transformer (25-50 words)

1.1 Average Treatment Effect (Per Experiment in Experiment Set)

cpu_hours	-0.270477 +/- 0.00199311 (stderr) (p=6.31722e-10, Welch's t-test)
gpu_hours	-0.240519 +/- 0.0019161 (stderr) (p=4.6659e-10, Welch's t-test)
estimated_carbon_impact_kg	-0.0246474 +/- 0.00204375 (stderr) (p=2.46912e-05, Welch's t-test)
total_power	-0.0672016 +/- 0.00649084 (stderr) (p=5.12301e-05, Welch's t-test)
kw_hr_gpu	-0.030893 +/- 0.000287623 (stderr) (p=1.01822e-09, Welch's t-test)
kw_hr_cpu	-0.0116397 +/- 0.00426262 (stderr) (p=0.00863053, Welch's t-test)
exp_len_hours	-0.268833 +/- 0.00199048 (stderr) (p=4.50651e-10, Welch's t-test)
average_realtime_carbon_intensity	-30.7896 +/- 9.31069 (stderr) (p=0.000821879, Welch's t-test)

1.2 Cumulative Difference (Per Experiment Set)

cpu_hours	-1.3523833333333335
gpu_hours	-1.2025934786925934
estimated_carbon_impact_kg	-0.12323679386487159
total_power	-0.33600820451063107
kw_hr_gpu	-0.15446508208424514
kw_hr_cpu	-0.058198338492103645
exp_len_hours	-1.344166325065825
average_realtime_carbon_intensity	-153.9478065954861