

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

1.1 Average Treatment Effect (Per Experiment in Experiment Set)

cpu_hours	0.0979972 +/- 0.00470954 (stderr) (p=6.93463e-08, Welch's t-test)
gpu_hours	0.0865428 +/- 0.00350667 (stderr) (p=1.68231e-09, Welch's t-test)
estimated_carbon_impact_kg	0.0128563 +/- 0.00205954 (stderr) (p=0.000297084, Welch's t-test)
total_power	0.0268959 +/- 0.00660061 (stderr) (p=0.00153114, Welch's t-test)
kw_hr_gpu	0.0113707 +/- 0.000475358 (stderr) (p=5.23346e-10, Welch's t-test)
kw_hr_cpu	0.00565204 +/- 0.00428206 (stderr) (p=0.080199, Welch's t-test)
exp_len_hours	0.0973839 +/- 0.00473878 (stderr) (p=8.2305e-08, Welch's t-test)
average_realtime_carbon_intensity	55.2955 +/- 10.9932 (stderr) (p=2.18349e-05, Welch's t-test)

1.2 Cumulative Difference (Per Experiment Set)

cpu_hours	0.4899861111111111
gpu_hours	0.43271381348640814
estimated_carbon_impact_kg	0.0642815931520341
total_power	0.1344793460180791
kw_hr_gpu	0.056853296157224315
kw_hr_cpu	0.028260213980800433
exp_len_hours	0.48691934830612604
average_realtime_carbon_intensity	276.477607226416