

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

1.1 Average Treatment Effect (Per Experiment in Experiment Set)

cpu_hours	-0.172479 +/- 0.004295 (stderr) (p=2.06675e-07, Welch's t-test)
gpu_hours	-0.153976 +/- 0.0029956 (stderr) (p=5.30351e-08, Welch's t-test)
estimated_carbon_impact_kg	-0.011791 +/- 0.000399598 (stderr) (p=3.69371e-10, Welch's t-test)
total_power	-0.0403058 +/- 0.00132659 (stderr) (p=7.69017e-08, Welch's t-test)
kw_hr_gpu	-0.0195224 +/- 0.000388133 (stderr) (p=5.01579e-08, Welch's t-test)
kw_hr_cpu	-0.00598762 +/- 0.000500355 (stderr) (p=1.67084e-06, Welch's t-test)
exp_len_hours	-0.171449 +/- 0.00433699 (stderr) (p=2.09888e-07, Welch's t-test)
average_realtime_carbon_intensity	24.506 +/- 8.99849 (stderr) (p=0.00220109, Welch's t-test)

1.2 Cumulative Difference (Per Experiment Set)

cpu_hours	-0.8623972222222224
gpu_hours	-0.7698796652061851
estimated_carbon_impact_kg	-0.05895520071283748
total_power	-0.201528858492552
kw_hr_gpu	-0.09761178592702084
kw_hr_cpu	-0.02993812451130321
exp_len_hours	-0.8572469767596987
average_realtime_carbon_intensity	122.52980063092991