Correlation Matrix: Benchmarks and Human-Model Alignment 1.00 0.44 0.67 0.43 0.03 0.50 0.68 0.64 0.67 ifeval 1.0 0.441.000.720.880.500.940.740.700.64bbh

- 0.8

- 0.6

0.670.721.000.890.190.840.610.570.51math

0.43 0.88 0.89 1.00 0.37 0.88 0.56 0.51 0.46 gpga

0.03 0.50 0.19 0.37 1.00 0.45 0.36 0.34 0.33 musr

0.500.940.840.880.451.000.710.680.62

mmlu

- 0.4 0.680.740.610.560.360.711.000.990.97-0.2

0.64 0.70 0.57 0.51 0.34 0.68 0.99 1.00 0.98

gpqa musr

rdm corr

pph

nath

0.67 0.64 0.51 0.46 0.33 0.62 <mark>0.97 0.98 1.00</mark>