

Model performance across maths and knowledge benchmarks

