

GPQA: A Graduate-Level Google- Proof Question and Answer Benchmark

