2. For analyses reporting issues (scores 0 or 0.5), generate m meta-verification assessments to validate the identified problems. An analysis is deemed valid if the majority of meta-assessments confirm its findings
3. For each proof, we examine analyses that assign the lowest score. If at least k such analyses are deemed valid, the proof is labeled with that lowest score. If no legitimate issues are identified across all verification attempts, the proof is labeled with 1. Otherwise, the proof is discarded or routed to human experts for labeling

In our last two training iterations, this fully automated pipeline replaced human annotation entirely. Quality checks confirmed that the automated labels aligned well with expert judgments.

3. Experiments

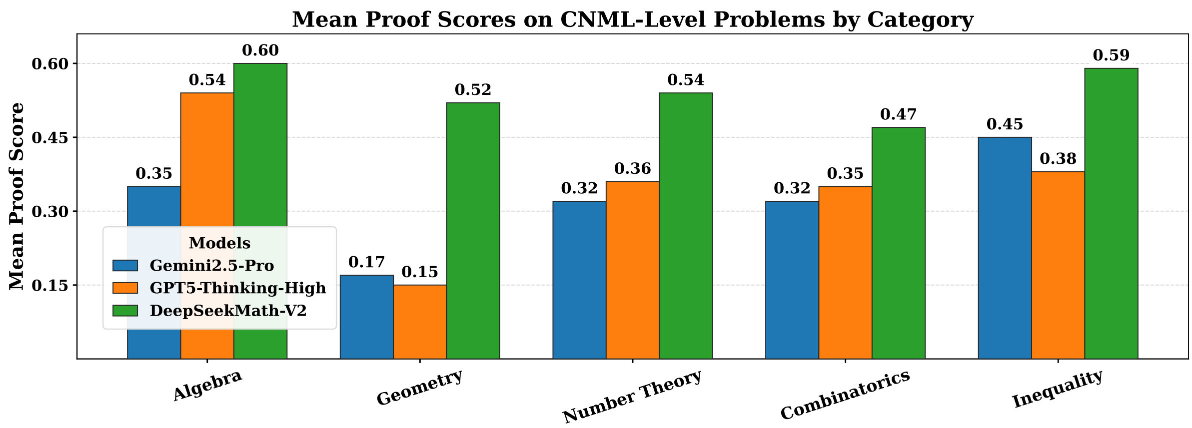

Figure 1 | Average proof scores on CNML-level problems by category and model, as evaluated by our verifier.

3.1. Training Settings

We employed Group Relative Policy Optimization (GRPO) (Shao et al., 2024) for reinforcement learning, iteratively optimizing proof verification and generation capabilities as described in Section 2. In each iteration, we first optimized proof verification. The proof generator was then initialized from the verifier checkpoint and optimized for proof generation. Starting from the second iteration, the proof verifier was initialized with a checkpoint that consolidated both verification and generation capabilities from the previous iteration through rejection fine-tuning.

3.2. Evaluation Benchmarks

We evaluate our final proof generator on the following theorem proving benchmarks:

In-House CNML-Level Problems 91 theorem-proving problems spanning algebra (13), geometry (24), number theory (19), combinatorics (24), and inequality (11), comparable in difficulty to problems from Chinese National High School Mathematics League (CNML)

Competition Problems

- **IMO 2025** (6 problems): The International Mathematical Olympiad, the premier global mathematics competition for pre-university students