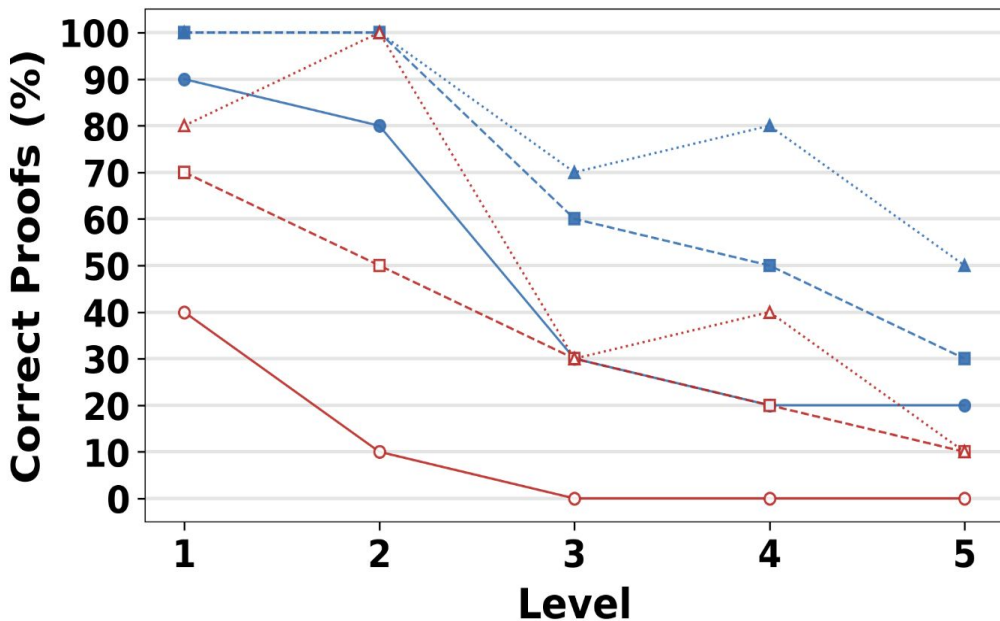


% correct proofs per level of difficulty



- Analogy-based - 1 run, no retries
- -■- Analogy-based - 1 run, 5 retries (w\ verifier)
- ...▲... Analogy-based - 3 runs, 5 retries (w\ verifier)
- Base model - 1 run, no retries
- -□- Base model - 1 run, 5 retries (w\ verifier)
- ...△... Base model - 3 runs, 5 retries (w\ verifier)