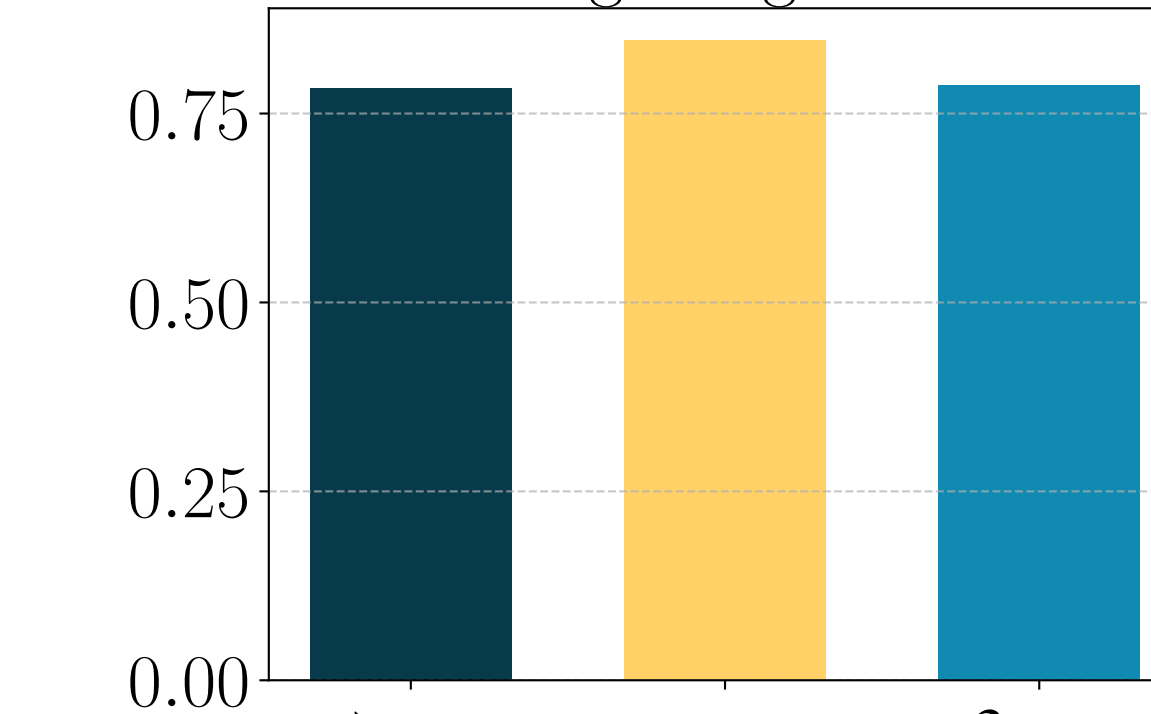
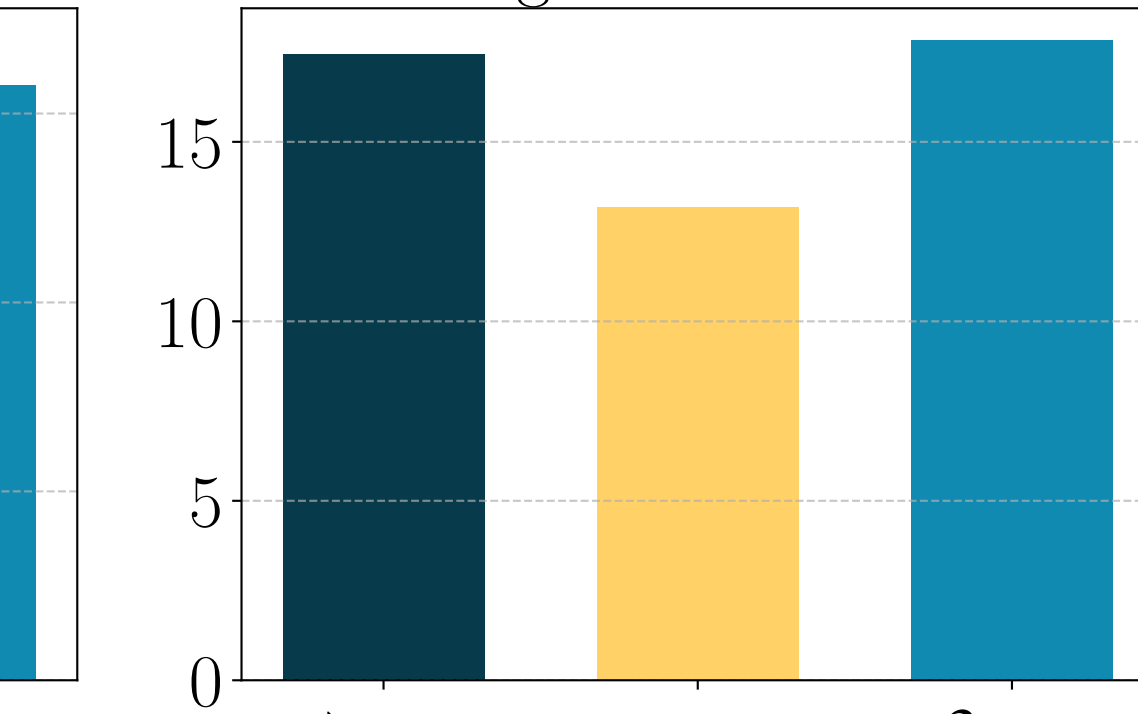


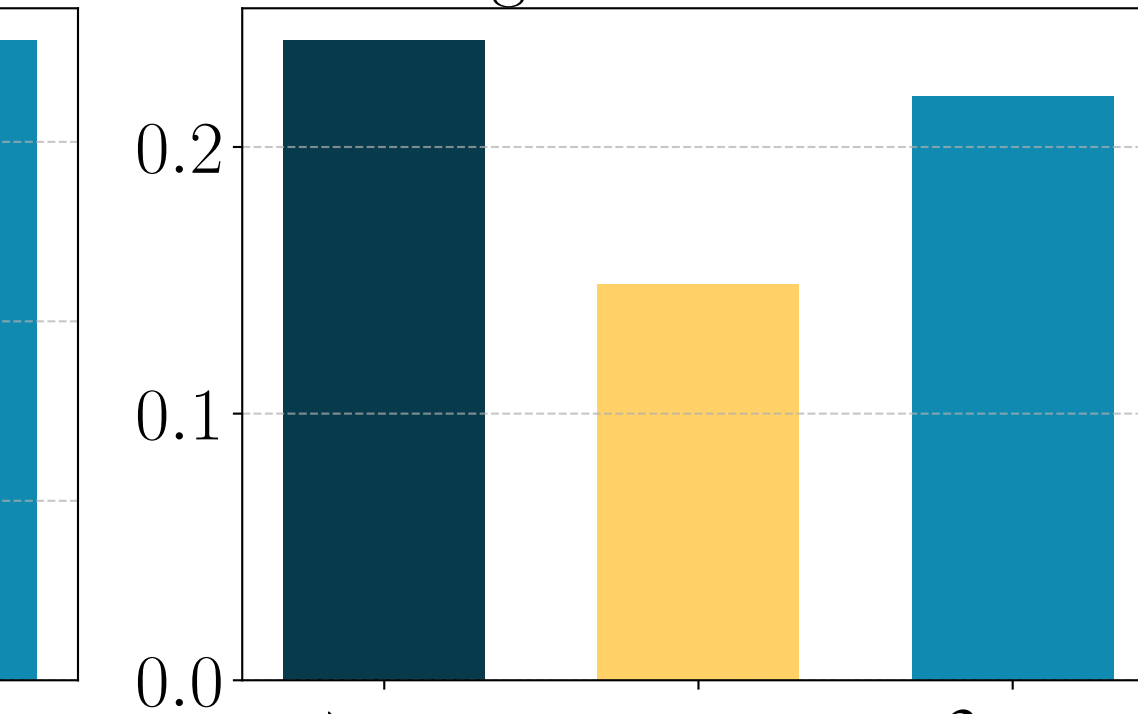
Forgetting Rate



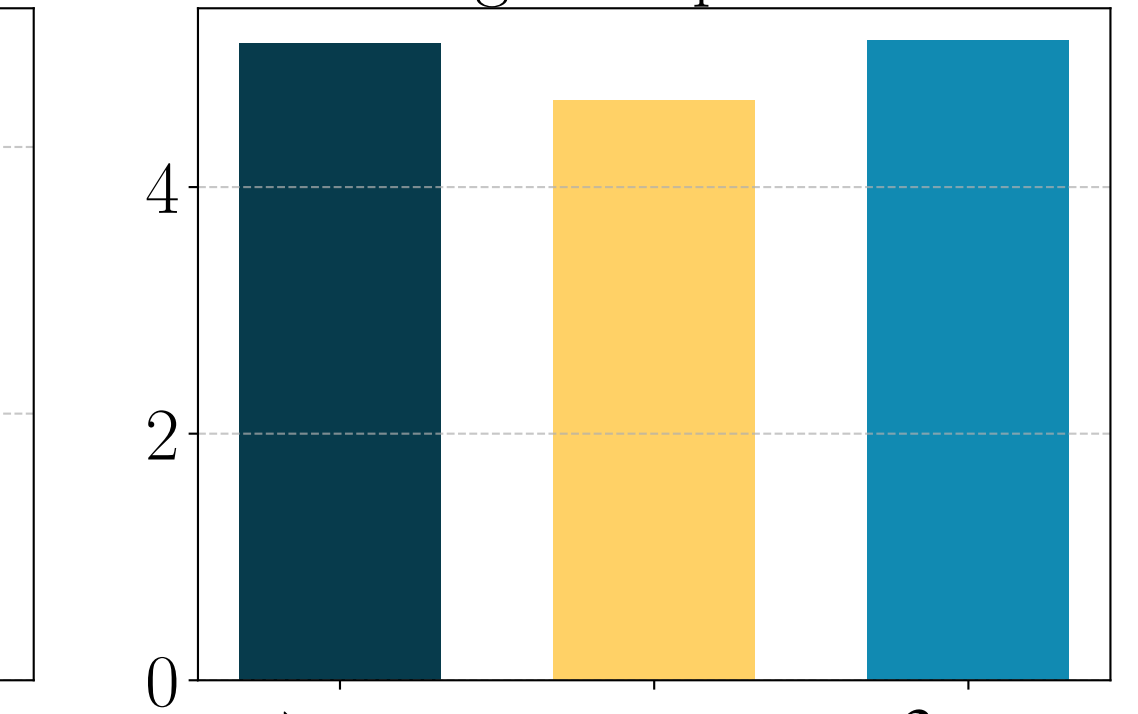
Average Solution Count



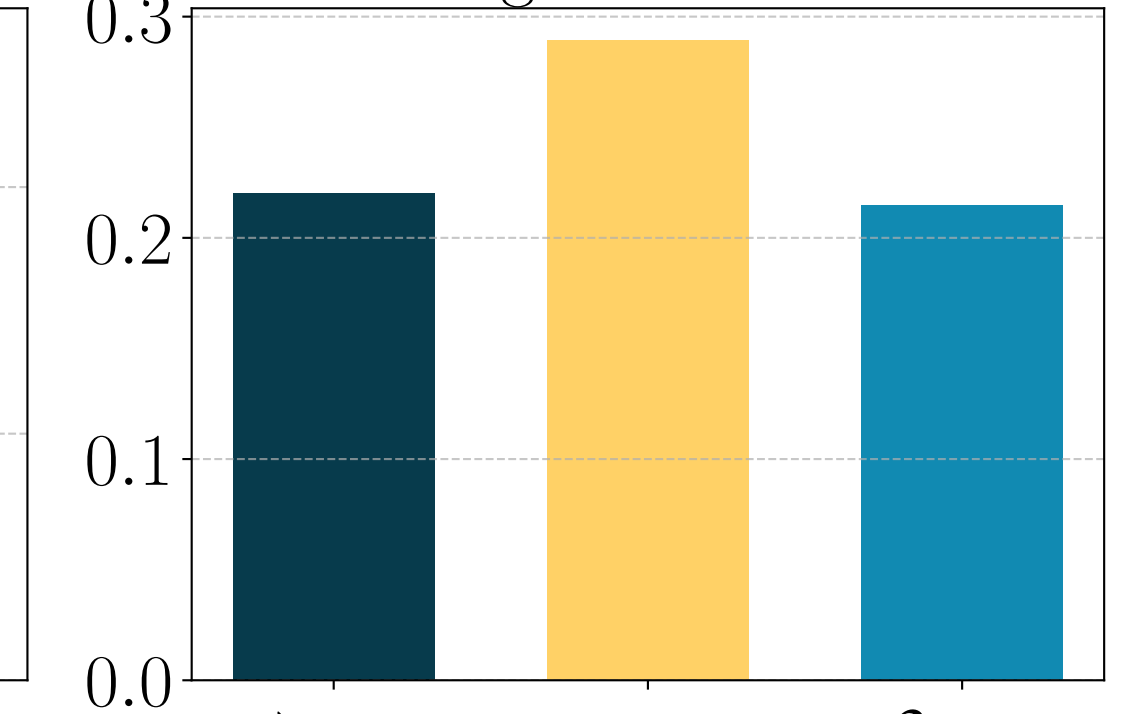
Average Verification Rate



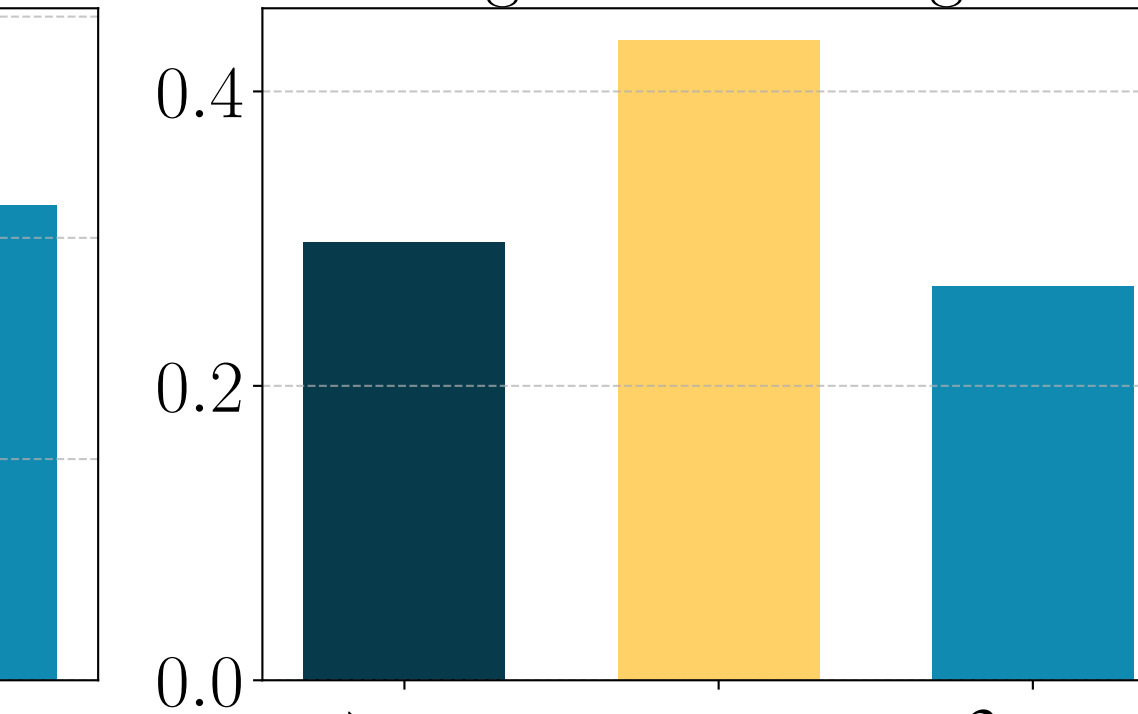
Average Jump Distance



Average Success Rate



Average Overthinking Rate



Pass@1

