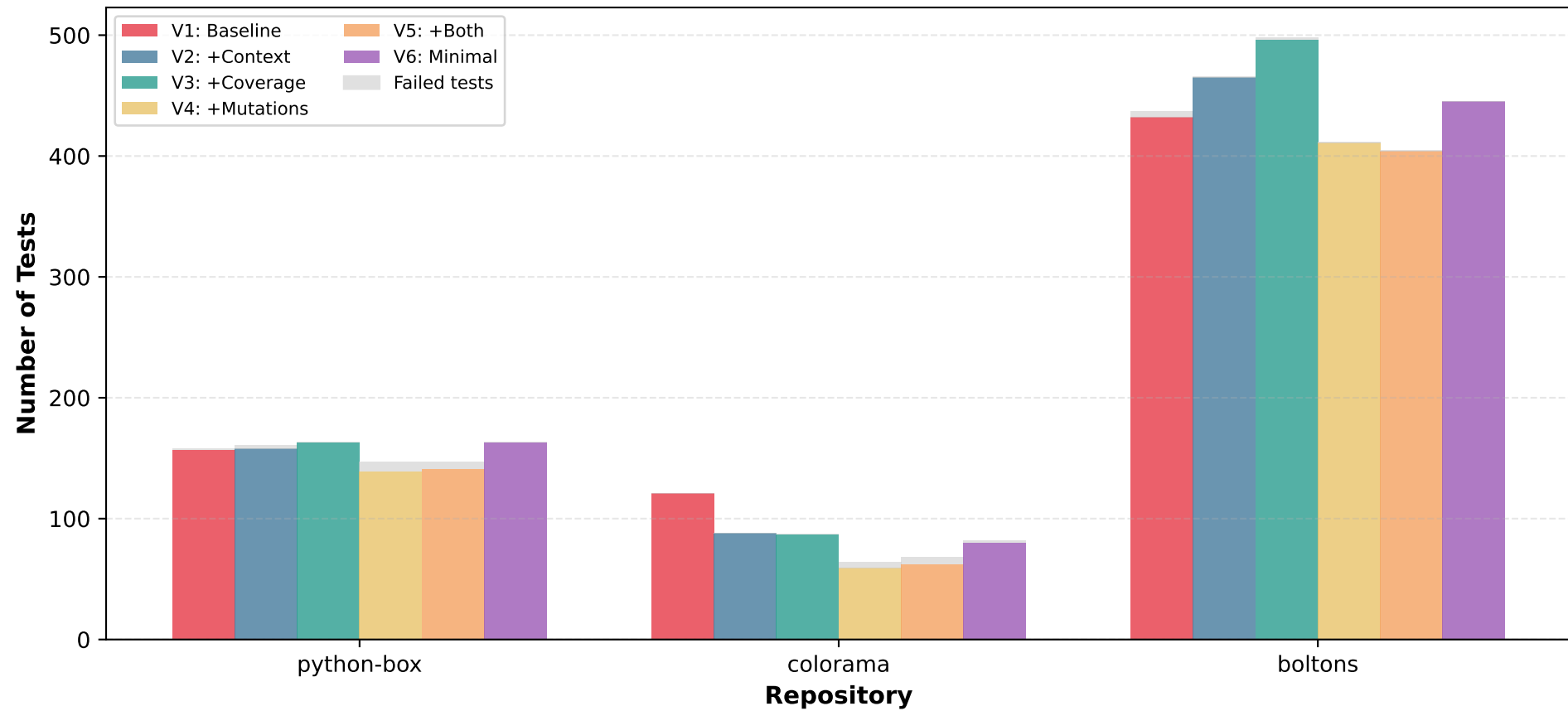
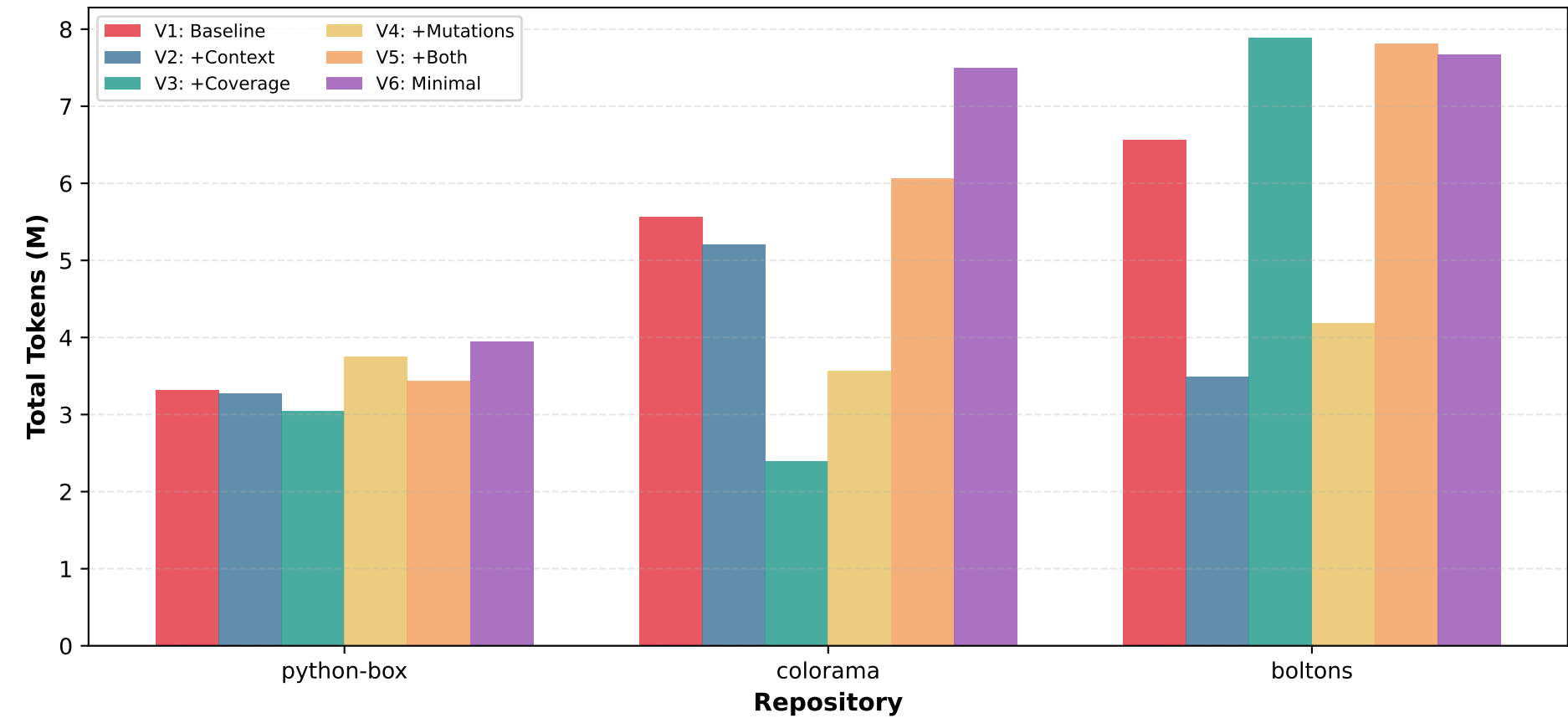


## Benchmark Results: 6 Prompt Variants Across 3 Repositories

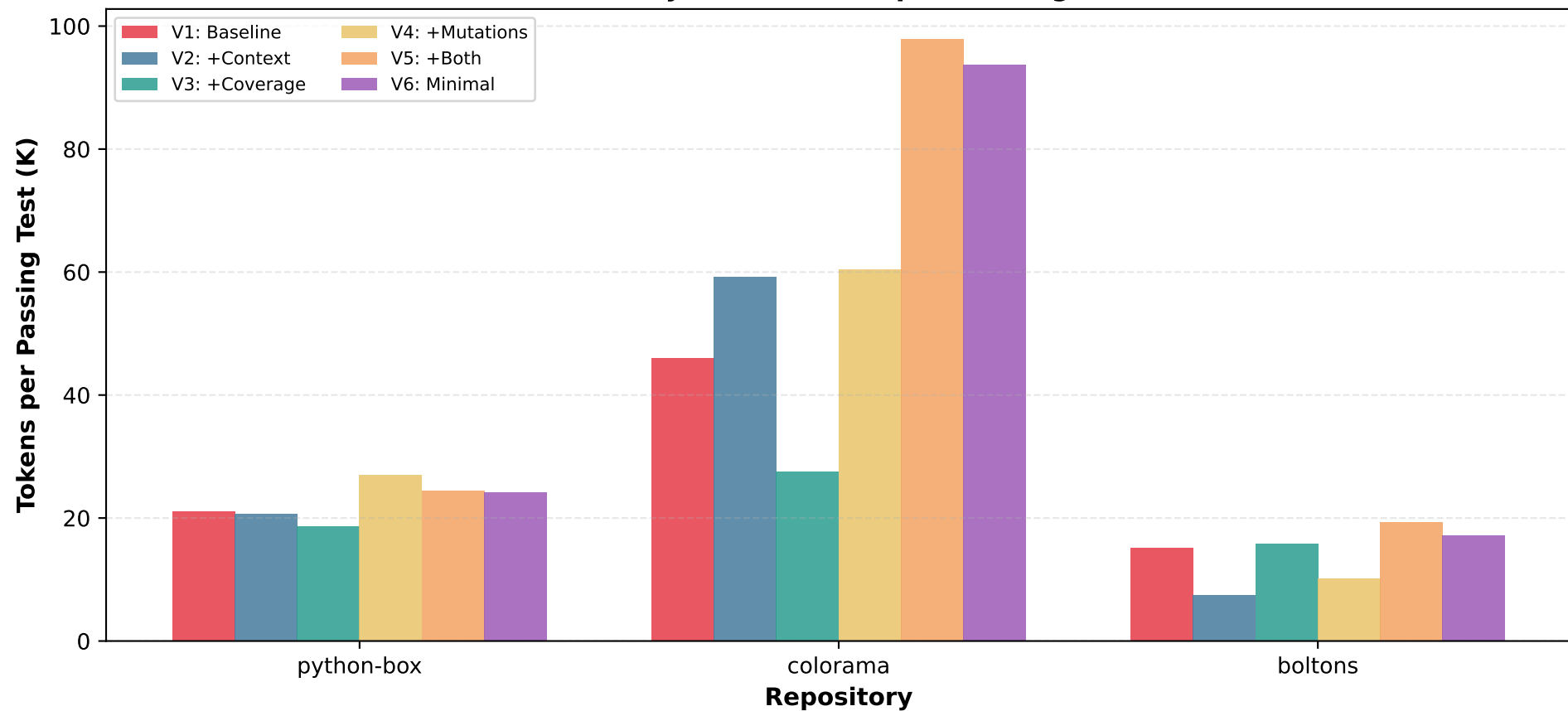
### Test Pass/Fail Rates by Variant and Repository



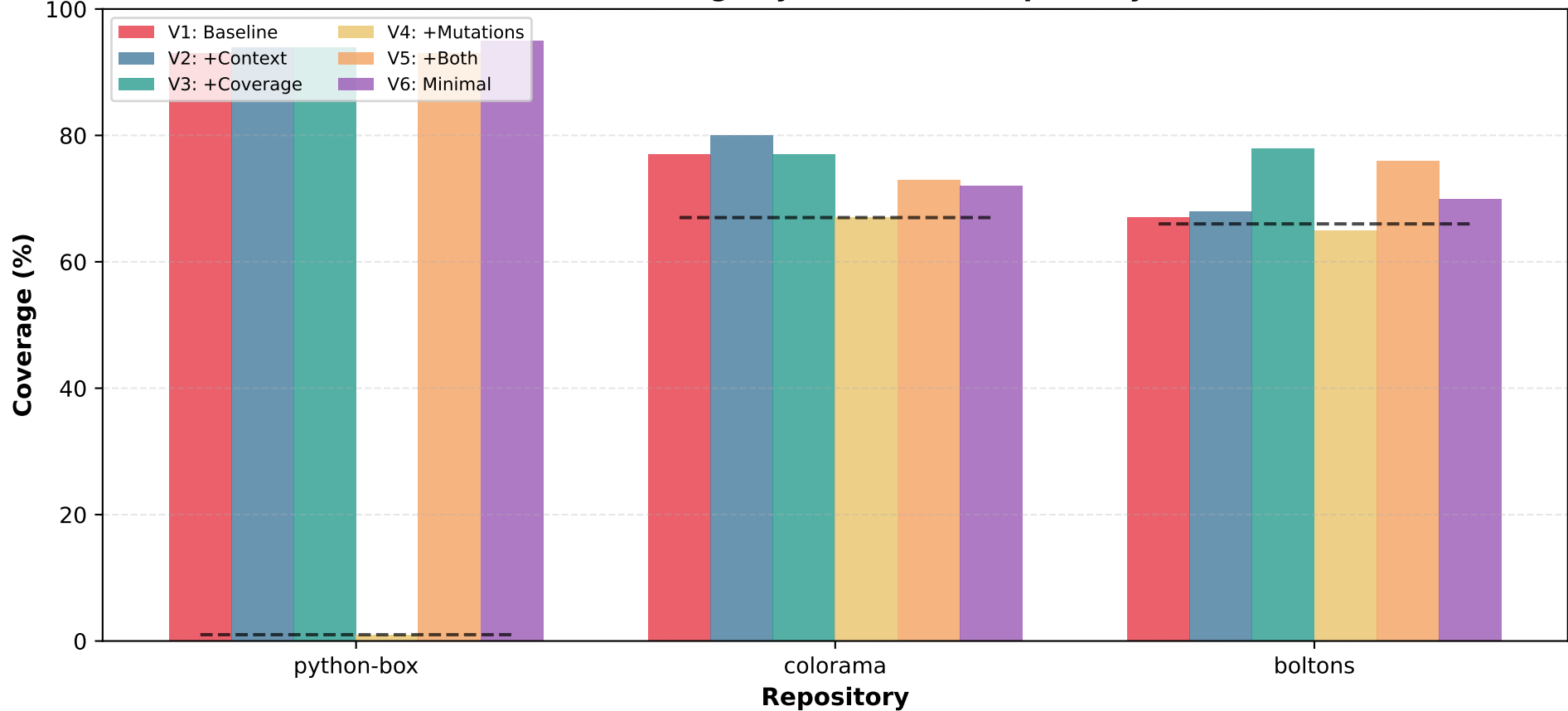
### Token Usage by Repository and Variant (Including Cache)



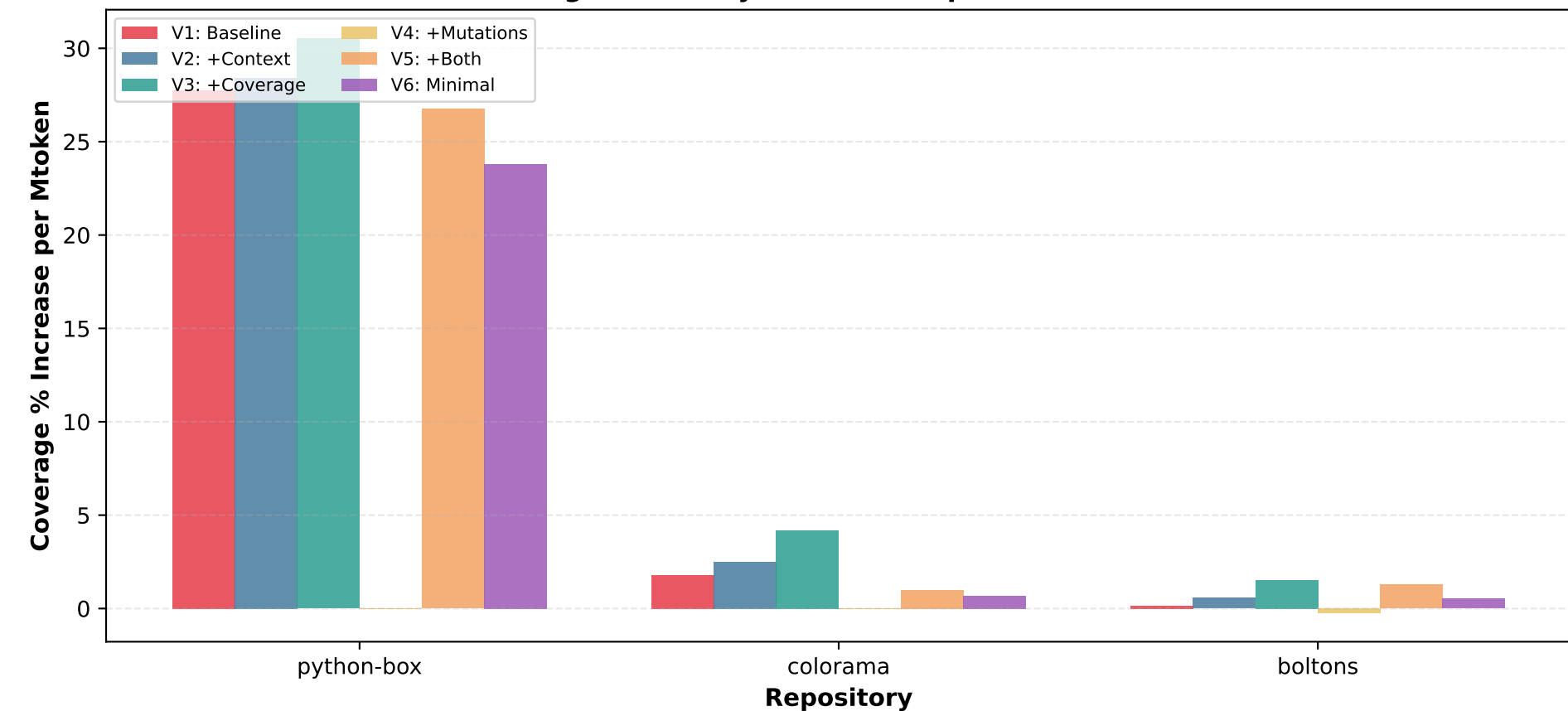
### Efficiency: Total Tokens per Passing Test



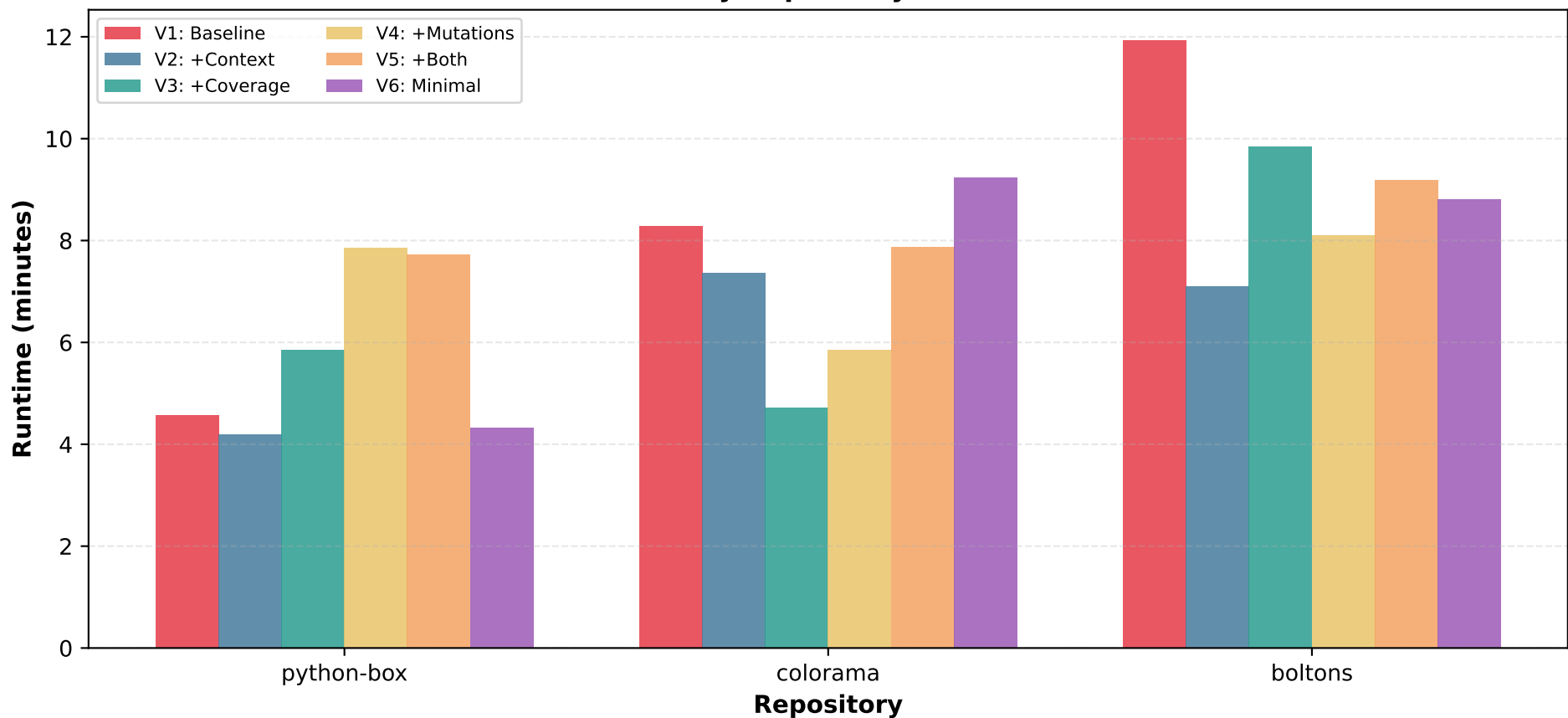
### Final Coverage by Variant and Repository



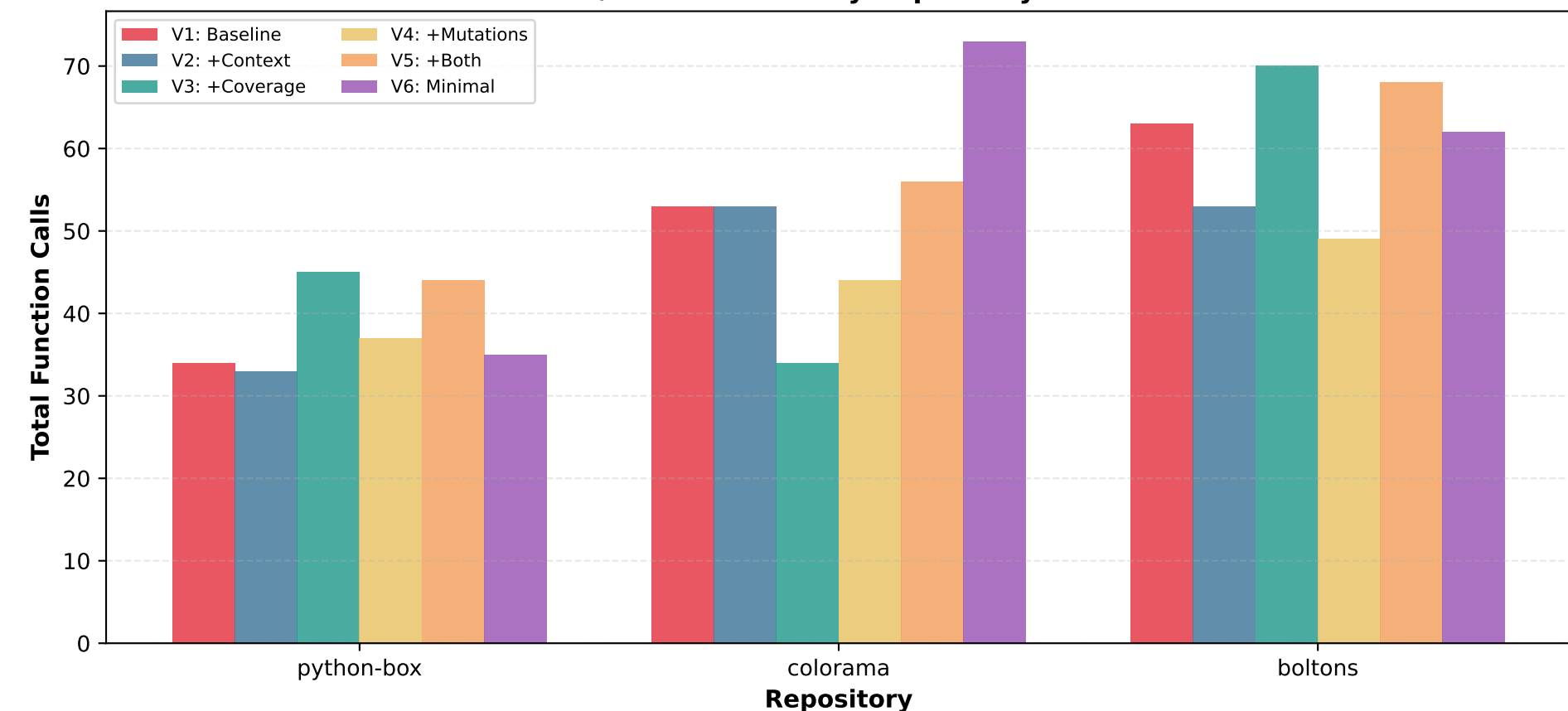
### Coverage Efficiency: % Increase per Million Tokens



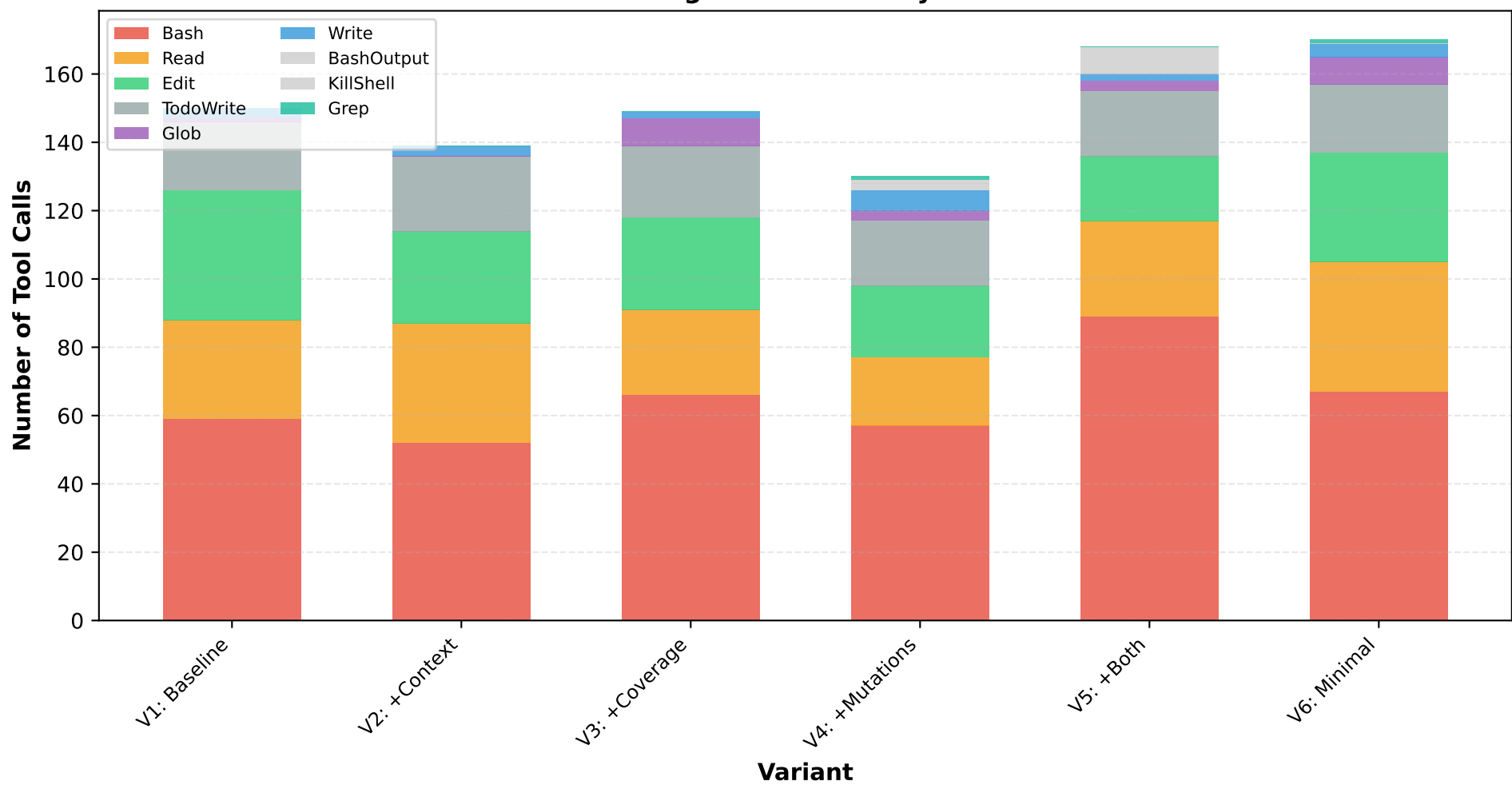
### Runtime by Repository and Variant



### Total Tool/Function Calls by Repository and Variant



### Tool Usage Breakdown by Variant



### Bash Command Type Breakdown by Variant

