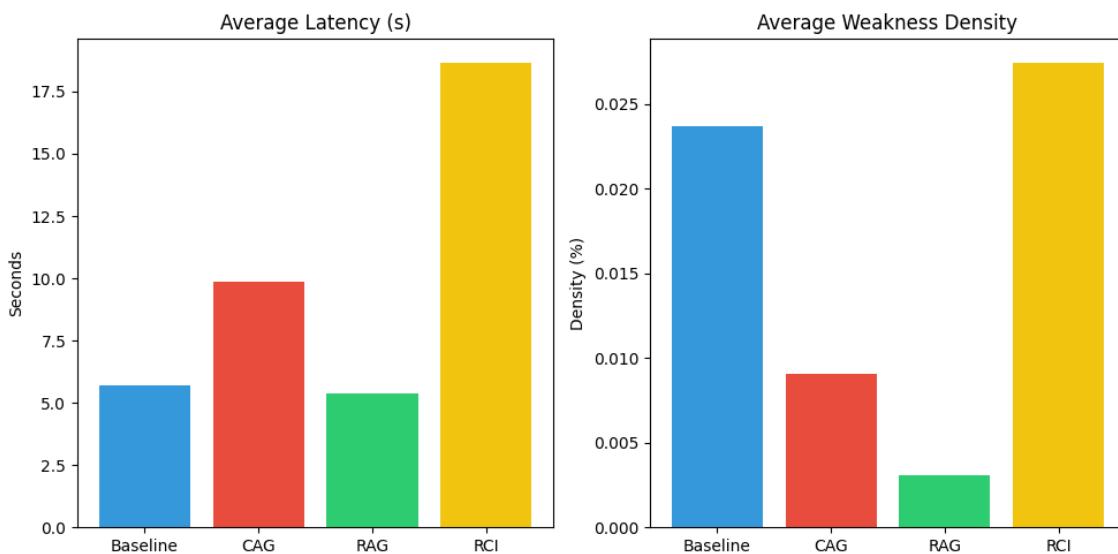


Security Evaluation Report

Automated Benchmarking of LLM Code Generation Techniques

1. Performance Visualization



2. Detailed Benchmark Data

Task	Approach	Latency	Weaknesses	Density
0	Baseline	5.68	1	1.10%
0	RAG	4.78	0	0.00%
0	RCI	21.84	3	2.59%
0	CAG	8.61	0	0.00%
1	Baseline	2.77	0	0.00%
1	RAG	6.59	0	0.00%
1	RCI	14.95	0	0.00%
1	CAG	5.13	0	0.00%
2	Baseline	7.28	4	3.88%
2	RAG	4.71	1	1.54%

2	RCI	21.25	4	2.67%
2	CAG	6.36	0	0.00%
3	Baseline	4.79	2	3.28%
3	RAG	3.93	0	0.00%
3	RCI	15.07	2	5.00%
3	CAG	13.08	2	4.55%
4	Baseline	7.89	4	3.60%
4	RAG	6.87	0	0.00%
4	RCI	20.26	3	3.49%
4	CAG	16.25	0	0.00%

3. Aggregated Scanner Findings

Bandit Analysis: Detected 27 total issues.

- B603: 9 occurrences
- B113: 7 occurrences
- B404: 6 occurrences

Semgrep Analysis: Detected 2 total issues.

Note: Individual findings are stored as JSON artifacts in the project directory.