



# Cross-Architecture Performance Analysis

This section provides per-architecture breakdown (sec:per-arch). For statistical significance, see sec:statistical-validation. For parameter sensitivity, see sec:sensitivity. For ablation studies and scalability, see sec:extended-ablation.

**Reproducibility:** All results generated by `scripts/run_full_evaluation.py` → `output/data/full_evaluation_results.json`.

**Implementation:** All evaluation infrastructure is in `src/evaluation/`. Key modules: `runner.py`: `ExperimentRunner` orchestrates  $950 \times 6$  evaluation matrices; `roc.py` computes ROC curves and AUC with bootstrap confidence intervals; `benchmark.py` measures latency and throughput; `metrics.py` computes detection rates, precision, recall, and F1 scores.

## Per-Architecture Breakdown

### Claude Code (Hierarchical Architecture)

#### Architecture Characteristics:

- ▶ Primary agent: Orchestrator with full context