RESEARCH SUMMARY / SEPTEMBER 17, 2026

Token reduction is not cost reduction

Corrected summary of version 5 of the PointFive paper. This replaces the earlier field guide and its generalized savings-ceiling claims.

What the experiment measured

Four Claude Code configurations, 103 tasks, seven repositories, and three models. The campaign executed 2,908 runs; 2,848 were included in the primary analysis.

Configuration	Delivered tool-output token change	Pooled billed-cost change
Unmodified Claude Code	Baseline	Baseline
RTK v0.44.1	-1.3%	-2.7%
RTK-ML, PointFive experimental build	-38.4%	+6.8%
Headroom v0.27.0	Not measured	+48.4%

These are paired billed-cost changes against the baseline, not cost per successful task. RTK's pooled estimate was below zero, but its holdout-only confidence interval crossed zero. The paper reports intervals, exclusions, and success metrics separately.

How to use the findings

Evaluate the complete task

Token reduction alone is insufficient evidence of a lower bill. Compare provider charges and task outcomes across representative workloads. Include retries, caching, additional model calls, and service fees. Hold the task and acceptance criteria constant and check whether the work still succeeds.

Keep the scope explicit

The results apply to the tested configurations. They do not establish a universal savings ceiling or describe every compression tool. RTK-ML was an experimental PointFive build, not unmodified RTK.

Disclosure: this is PointFive-authored research, not an independent evaluation of PointFive's products. This summary does not establish a product savings guarantee.

Sources and reproducibility

Version 5 paper and methodology
https://arxiv.org/abs/2607.12161v5

Results and confidence intervals, Table 8
https://arxiv.org/html/2607.12161v5#S7.T8

AI Efficiency Benchmark
https://github.com/PointFiveLabs/ai-efficiency-benchmark

For future use, check the versioned source and current provider pricing. Workload, model, caching behavior, and success criteria can change the result.