Comparative Governance Effects in a GPU Kernel Marketplace with Correlated Proxy Gaps

Raeli Savitt

February 2026

Abstract

We compare the effects of seven governance regimes on welfare, toxicity, and agent payoffs in a GPU kernel marketplace with correlated speedup-cheating, split functional/OOD test regimes, and tolerance exploitation. Sweeping from no governance to full governance (audits + staking + circuit breaker + reputation + tax) across 70 runs (7 regimes × 10 seeds), we find that **circuit breakers** emerge as the most effective single lever: they produce the highest welfare (22.96 vs 12.70 no-governance baseline, $d = 1.64$, $p = 0.022$ Bonferroni-corrected) and the lowest toxicity (0.395 vs 0.446, $d = -2.08$, $p = 0.003$ Bonferroni-corrected). Full governance also significantly reduces toxicity ($d = -1.89$, $p = 0.007$). By contrast, staking alone produces the worst outcomes on both welfare (10.65) and toxicity (0.452). Nine out of 42 pairwise comparisons survive Bonferroni correction, making this the most statistically productive sweep in the kernel market series. Agent-type stratification remains massive across all regimes (honest vs adversarial $d = 2.24$–3.39, all $p < 0.0001$), and quality gap stays positive (0.015–0.116), indicating the market avoids adverse selection under all tested governance configurations.

1 Introduction

Previous work on the kernel market v2 model established two findings: (1) transaction taxes reduce welfare without reducing toxicity (deadweight loss, $d = 1.19$), and (2) audit rate primarily affects deterrence (adversarial payoffs) rather than system-level outcomes. Both studies held most governance levers constant, varying one at a time.

This paper takes a different approach: we compare **complete governance regimes** against each other, including a no-governance baseline. This allows us to answer: which governance mechanism matters most? Do combinations outperform single levers? And which mechanisms have no effect or negative effects?

We test seven regimes spanning the space from no governance to full governance:

1. **No governance**: All mechanisms disabled, reputation decay = 1.0
2. **Audits only**: 20% audit probability, 3× penalty multiplier
3. **Staking only**: Minimum stake = 1.0, slash rate = 0.3
4. **Reputation only**: Reputation decay = 0.85
5. **Circuit breaker only**: Freeze threshold toxicity = 0.5, 2 violations, 3-epoch duration
6. **Audits + staking**: Both mechanisms enabled
7. **Full governance**: All levers enabled including 5% transaction tax

2 Model

We use the kernel market v2 scenario with correlated speedup-cheating, split functional/OOD test regimes, and tolerance exploitation (see companion v2 paper).

2.1 Agent Composition

Table 1: Agent composition in the v2 kernel market scenario.

Type	Count	Func. Quality	OOD Quality	Base Speedup	Role
Honest	2	0.85	0.80	$\mathcal{N}(0.50, \sigma)$	Correct + fast kernels
Opportunistic	3	0.70	0.50	$\mathcal{N}(0.40, \sigma)$	Cut corners on precision
Adversarial	2	0.95	0.30	$\mathcal{N}(0.45, \sigma) + 0.40$	Exploit proxy signals
Verifier	1	—	—	—	Surface-level checks

Speedup variance σ scales with difficulty: easy = 0.15, medium = 0.25, hard = 0.35.

2.2 Governance Regimes

Table 2: Governance regime configurations.

Regime	Audit Prob.	Staking	Rep. Decay	Circuit Breaker	Tax
No governance	0.00	Off	1.00	Off	0%
Audits only	0.20	Off	0.98	Off	0%
Staking only	0.00	On	0.98	Off	0%
Reputation only	0.00	Off	0.85	Off	0%
Circuit breaker only	0.00	Off	0.98	On	0%
Audits + staking	0.20	On	0.98	Off	0%
Full governance	0.15	On	0.90	On	5%

2.3 Sweep Configuration

Table 3: Parameter sweep design.

Parameter	Values
`governance.regime`	7 discrete regimes (Table 2)
Seeds per configuration	10
Total runs	**70**
Epochs per run	5
Steps per epoch	10

3 Results

3.1 Welfare

Table 4: Total welfare by governance regime.

Regime	Welfare (mean ± SD)
No governance	12.70 ± 5.28
Audits only	15.02 ± 4.17
Staking only	10.65 ± 4.13
Reputation only	15.18 ± 3.40
Circuit breaker only	$\mathbf{22.96 \pm 6.22}$
Audits + staking	13.34 ± 5.26
Full governance	21.38 ± 7.30

Circuit breaker achieves nearly double the welfare of staking (22.96 vs 10.65, $d = 2.20$, Bonferroni $p = 0.008$). It also significantly outperforms no governance ($d = 1.64$, $p = 0.022$ Bonferroni-corrected across 12 baseline comparisons).

3.2 Toxicity

Table 5: Toxicity rate by governance regime.

Regime	Toxicity (mean ± SD)
No governance	0.4463 ± 0.0282
Audits only	0.4319 ± 0.0178
Staking only	0.4518 ± 0.0224
Reputation only	0.4357 ± 0.0233
Circuit breaker only	$\mathbf{0.3948 \pm 0.0209}$
Audits + staking	0.4396 ± 0.0222
Full governance	$\mathbf{0.3992 \pm 0.0211}$

Circuit breaker and full governance are the only regimes that significantly reduce toxicity vs no governance:

- Circuit breaker: $d = -2.08$, Bonferroni $p = 0.003$
- Full governance: $d = -1.89$, Bonferroni $p = 0.007$

3.3 Quality Gap

Table 6: Quality gap by governance regime.

Regime	Quality Gap (mean)
No governance	0.113
Audits only	0.116
Staking only	0.099
Reputation only	0.113
Circuit breaker only	0.038
Audits + staking	0.097
Full governance	0.015

Quality gap remains positive across all regimes (no adverse selection). Circuit breaker and full governance have the lowest quality gaps, suggesting these regimes narrow the distribution of accepted submissions.

3.4 Agent-Type Stratification

Table 7: Agent payoffs by type and governance regime (paired t-test, honest vs adversarial).

Regime	Honest	Opportunistic	Adversarial	d (H vs A)
No governance	2.64	2.53	-1.40	2.27***
Audits only	3.28	2.93	-1.80	2.91***
Staking only	2.26	2.27	-1.46	3.39***
Reputation only	3.09	2.82	-1.28	2.86***
Circuit breaker only	4.30	3.74	-0.59	2.62***
Audits + staking	2.95	2.60	-1.66	2.53***
Full governance	4.34	3.40	-0.92	2.24***

***$p < 0.0001$, Bonferroni-significant.

Pooled effect sizes across all regimes:

- Honest vs adversarial: $d = 3.34$, $p < 0.00001$
- Opportunistic vs adversarial: $d = 5.42$, $p < 0.00001$
- Honest vs opportunistic: $d = 0.25$, $p = 0.14$ (ns)

3.5 Bonferroni-Significant Comparisons

Table 8: All Bonferroni-significant pairwise comparisons (9/42, $\alpha = 0.05/42 = 0.00119$).

Rank	Comparison	$\|d\|$	Bonf. p
1	Toxicity: CB vs Staking	2.64	0.0006
2	Toxicity: Full gov. vs Staking	2.42	0.0016
3	Welfare: CB vs Staking	2.20	0.008
4	Toxicity: Audits+staking vs CB	2.08	0.008
5	Toxicity: CB vs No governance	2.08	0.010
6	Toxicity: Audits only vs CB	1.91	0.020
7	Toxicity: Full gov. vs No gov.	1.89	0.024
8	Toxicity: Audits+staking vs Full gov.	1.87	0.024
9	Toxicity: CB vs Reputation	1.85	0.026

CB = Circuit breaker only. Seven of nine significant results involve circuit breakers.

3.6 Omnibus Tests

Table 9: Omnibus tests across all regimes.

Test	Metric	Statistic	p
Kruskal-Wallis	Welfare	$H = 24.24$	0.0005
Kruskal-Wallis	Toxicity	$H = 33.37$	9×10^{-6}
Kruskal-Wallis	Quality gap	$H = 36.29$	2×10^{-6}
ANOVA	Welfare	$F = 6.70$	1.6×10^{-5}
ANOVA	Toxicity	$F = 10.06$	$< 10^{-6}$
ANOVA	Quality gap	$F = 12.89$	$< 10^{-6}$

3.7 Normality Validation

Shapiro-Wilk tests confirm welfare and toxicity distributions are normal across all regime groups (all $p > 0.05$), validating the use of parametric t-tests. Mann-Whitney U tests confirm all Bonferroni-significant results.

4 Discussion

4.1 Circuit Breakers Dominate

The circuit breaker emerges as the single most effective governance mechanism, producing the highest welfare (+81% vs no governance), lowest toxicity (−11%), and the highest honest agent payoffs (+63%). Seven of nine Bonferroni-significant comparisons involve circuit breakers.

Paradoxically, circuit breakers are also the regime under which adversarial agents lose the *least* (-0.59 vs -1.40 no governance). The explanation: circuit breakers freeze high-toxicity agents, preventing them from further harming the ecosystem. This benefits everyone—including adversarial agents, who accumulate fewer penalties during frozen periods.

4.2 Staking Backfires

Staking alone produces the worst outcomes across the board: lowest welfare (10.65), highest toxicity (0.452), and the lowest honest/opportunistic payoffs. In correlated proxy-gap environments, requiring agents to post collateral before participating creates a barrier that disproportionately harms honest agents while failing to deter adversarial agents who can still recoup stake through inflated speedup signals.

4.3 The Tax Penalty Persists

Full governance (which includes a 5% transaction tax) achieves significantly lower toxicity than no governance ($d = -1.89$) but does not significantly outperform circuit breaker alone on any metric. The tax component appears to drag down welfare relative to what circuit breakers alone achieve (21.38 vs 22.96), consistent with the v2 finding that taxes impose deadweight loss.

4.4 Combinations Can Be Subadditive

Audits + staking performs worse than audits alone on welfare (13.34 vs 15.02) and toxicity (0.440 vs 0.432). Adding staking to audits dilutes the audit benefit rather than amplifying it. This suggests governance interactions can be **subadditive**—the whole is less than the sum of parts.

4.5 Honest ≈ Opportunistic Persists

As in the audit rate study, honest and opportunistic agents earn nearly identical payoffs ($d = 0.25$, ns). The proxy gap between these types is not large enough for governance to distinguish them.

5 Reproducibility

```
# See scenarios/kernel_market/sweeps/governance_comparison.yaml
# for exact regime definitions.
# Run with: python examples/parameter_sweep.py \
#   --scenario scenarios/kernel_market/v2.yaml \
#   --sweep-config scenarios/kernel_market/sweeps/governance_comparison.yaml \
#   --seeds 10
```

Run artifacts: `runs/20260211-000149_kernel_market_governance_comparison/`

6 Figures

1. `plots/welfare_by_regime.png` — Mean welfare by regime with 95% CI
2. `plots/toxicity_by_regime.png` — Mean toxicity by regime with 95% CI
3. `plots/quality_gap_by_regime.png` — Quality gap with adverse selection threshold
4. `plots/agent_payoff_by_regime.png` — Agent payoffs by type and regime
5. `plots/welfare_toxicity_tradeoff.png` — Welfare-toxicity scatter
6. `plots/adversarial_payoff_by_regime.png` — Adversarial payoff with significance
7. `plots/regime_heatmap.png` — Z-scored performance heatmap

7 Limitations

- **5 epochs per run** may be insufficient for circuit breaker dynamics to fully manifest (freezes are 3 epochs long).
- **Fixed agent composition**—future sweeps should vary the adversarial fraction.
- **No interaction sweeps**—we test 7 point configs, not the full parameter space.
- **Non-adaptive agents**—agents don't change strategy in response to governance.

8 Conclusion

Comparing seven governance regimes in the v2 kernel market model reveals that **circuit breakers are the dominant governance mechanism**, producing the highest welfare (+81% vs no governance), lowest toxicity (−11%), and the most Bonferroni-significant comparisons (7/9 involve circuit breakers). Staking alone backfires, reducing welfare and increasing toxicity. Transaction taxes impose deadweight loss even within full governance. Governance combinations can be subadditive (audits + staking < audits alone). These findings suggest that **mechanism design matters more than mechanism quantity**: a single well-targeted lever (circuit breaker) outperforms a full governance stack that includes less effective mechanisms.