Against the OWASP Core Rule Set, and after learning

The same scored requests put through Senua AI and through the OWASP Core Rule Set on ModSecurity v3, exactly as it ships. Nothing tuned on either side.

2.4x the attacking requests blocked by the rule set, once the engine has learned from confirmed attacks, at the same false-block rate of 2 in 18,000

RESULTS

MEASURE	SENUA AI	RULE SET
Attacking sources stopped, untuned	66.87%	64.71%
Genuine visitors wrongly blocked, untuned	1 in 18,000	2 in 18,000
Attacking requests blocked, untuned	29.83%	28.03%
Attacking requests blocked, after learning	**67.40%**	**28.03%**
Genuine wrongly blocked, after learning	2 in 18,000	2 in 18,000

HOW IT WAS MEASURED

Baseline	owasp/modsecurity-crs:nginx, stock, paranoia level 1, anomaly threshold 5
Senua AI, untuned	Learned normal traffic only; level chosen on separate data at the rule set's own false-block rate
Senua AI, after learning	Confirmed attacks from a feedback half recorded; scored on the other half, never fed back
Significance	McNemar's test on the paired outcomes of the same requests

CONDITIONS

Untuned lead, significance	McNemar chi-squared 19.14, p = 1.2e-05
Attacks caught only by Senua AI	1,423
Attacks caught only by the rule set	1,198
Rule set turned up one level	blocks every genuine visitor

LIMITS OF THIS RESULT

CSIC attacks are generated from templates, so the after-learning result shows the engine recognising new instances of attack types it has been shown examples of, not anticipating attack families it has never seen in any form.

WHERE IT IS RECORDED

simplex/validate/BENCHMARKING.md (B2, B3/B4); simplex/validate/csic2010/out/b2_summary.json. The data behind this sheet is published beside it at senuaai.com/evidence/datasheets/.