Gauntlet

The Security Layer for AI

"Proof of Classifier Intelligence Through Pressure."

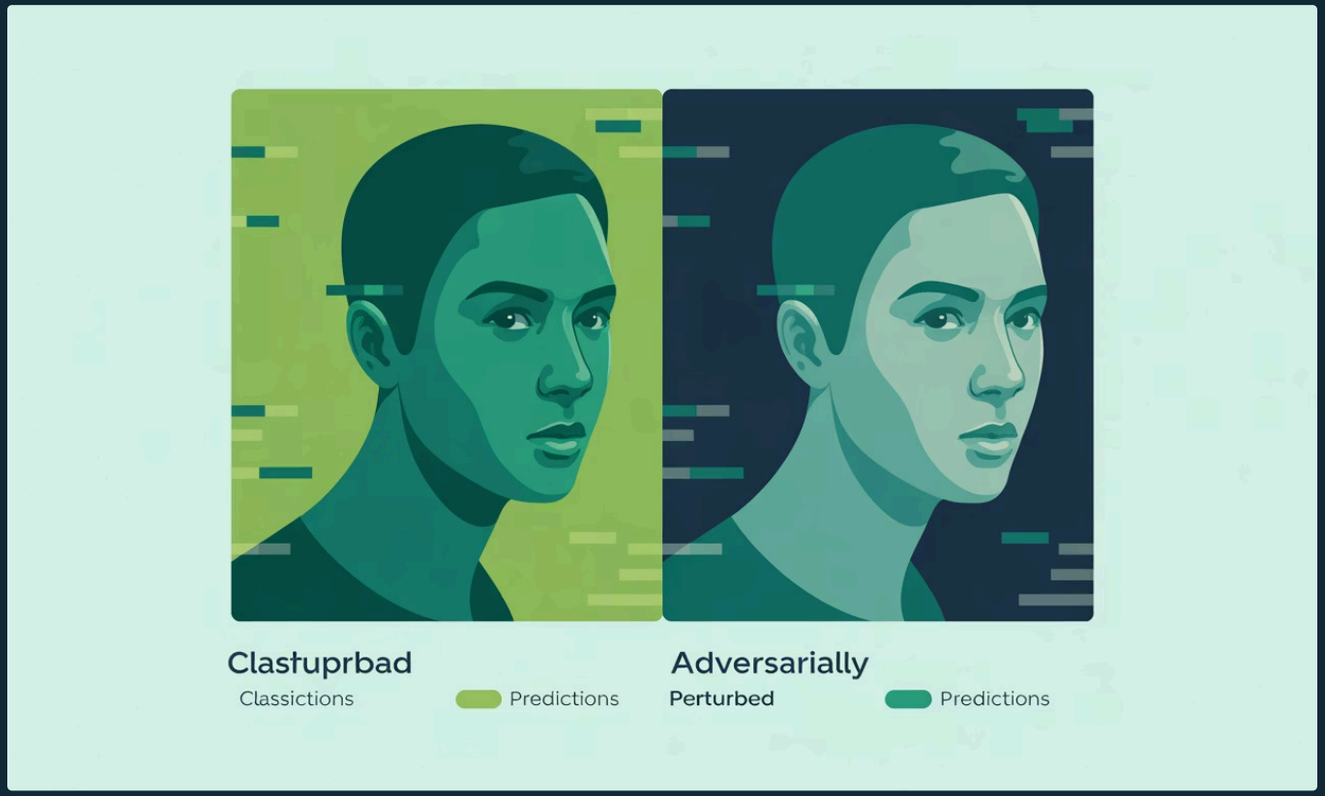
BITTENSOR SUBNET PROPOSAL

The Problem

AI Is Brittle

Most AI models fail under minimal adversarial pressure.

- Tiny perturbations can flip model predictions
- AI is increasingly deployed in high-stakes environments
- Security testing is centralized, static, and expensive
- No continuous robustness benchmark exists



The Opportunity

AI Security Market

Robust AI will become mandatory infrastructure.

AI adoption in finance, healthcare, defense, and autonomous systems

Regulatory pressure for AI reliability and safety

Growing enterprise demand for red-teaming and robustness validation

No decentralized robustness oracle exists

Robustness will become a measurable, monetizable signal.

The Solution

Gauntlet Subnet

A decentralized adversarial robustness arena.

Miners host classifiers

Models are deployed and defended by miners competing for emissions.

Validators generate adaptive adversarial attacks

Attack engines probe for weaknesses in real time.

Emissions reward accuracy under attack

Only models that survive earn rewards.

Continuous competitive pressure improves models

An ever-escalating arms race drives robustness forward.

The Gauntlet loop ensures that only the most resilient classifiers survive and earn.

How It Works

Mechanism Design

Incentivized adversarial competition.

Scoring Formula

$$Score = \alpha \cdot \text{Robust Accuracy} + \beta \cdot \text{Clean Accuracy} - \text{Latency Penalty}$$

$$Emission \propto Score^{\tau}$$

Key Mechanics

- Robust accuracy weighted highest
- Validators rewarded for exposing weaknesses
- Hidden datasets prevent overfitting
- Temperature parameter sharpens competition

Only resilient models earn emissions.

Why This Is

Proof of Intelligence

Surviving attack requires real intelligence.

Adversarial training required

Models must actively learn to defend against sophisticated attacks.

Resists adaptive gradient-based attacks

True robustness means surviving the strongest known attack methods.

Penalizes gradient masking tricks

Fake robustness through obfuscation is detected and punished.

Continuous pressure prevents stagnation

The arena never stops evolving — neither can the models.

Intelligence that survives pressure is intelligence that matters.

Competitive Landscape

Competitive Landscape

No live decentralized robustness market exists.

Outside Bittensor

- Academic benchmarks (static datasets)
- Internal red-teaming (centralized)
- RobustBench-style leaderboards

Within Bittensor

- Inference subnets reward accuracy
- No subnet focused on adversarial resilience

Gauntlet is the first decentralized robustness arena.

Business Model

Business Model & Long-Term Value

Robustness as an on-chain security primitive.

Enterprise robustness certification

Verified resilience scores for production AI systems.

API access to live robustness scores

Real-time robustness data for integration into any pipeline.

AI insurance underwriting inputs

Quantified risk signals for AI liability coverage.

White-label adversarial testing

Turnkey robustness testing for enterprise clients.

Long-term vision: On-chain robustness oracle for AI systems.

Go-To-Market

Go-To-Market Strategy

Bootstrapping a competitive intelligence arena.

Initial Target Users

- AI startups deploying classifiers
- Web3 AI protocols
- Security-focused research labs

Early Incentives

- Bonus emissions for early miners
- Bounties for validators breaking top models
- Public leaderboard visibility

Distribution

- Crypto-native AI communities
- Bittensor ecosystem
- Research & hackathon exposure

Vision

The Future of AI Security

Gauntlet becomes the resilience layer of AI.

01

Continuous adversarial benchmarking

02

Multimodal robustness expansion

03

LLM jailbreak resistance

04

AI risk scoring infrastructure

As AI systems become critical infrastructure, resilience will matter more than raw intelligence.

Run the Gauntlet.