# Validating Spam and Scam Detection

## Red Team Testing as an Alternative to Honeypots for Validating AI-Powered Call Analytics Engines.

Tuesday, September 16

# Using AI to Validate System Performance

Analytic systems continue to gain complexity paralleling the growth in Artificial Intelligence (AI)

# How Do We Test Performance Today?

**Today we use several methods to test the performance of an analytics solution.**

**Honeypots –**
**Static and dynamic**

**A/B Testing**

**PEN Testing**

# What Tools Can We Use?

Bad actors will use current tools

# Bad Actors

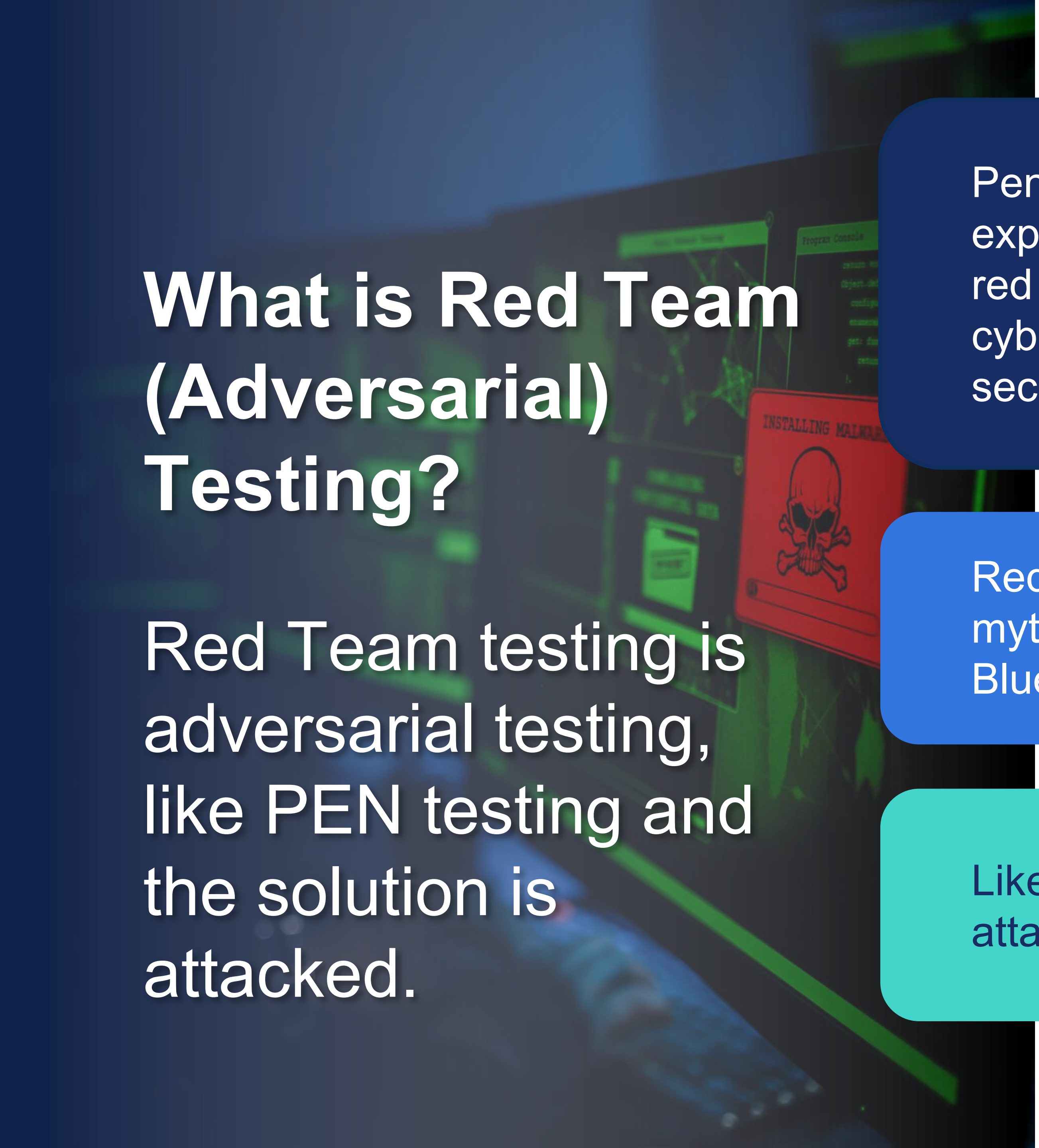
- Deeper Use of LLMs
- Voice Cloning
- Video Cloning
- Dark Web Data Mining
- Botnets and Malware for eSIM/Account Takeovers
- Enabling Remote SIM Swaps
- Fake eSIM Profiles
- Bypass of STIR/SHAKEN

**What are bad actors doing?**

# Can bad actors circumvent conventional testing methods?

# Summary of Bad actors in Action

**We can do the same!**



# What is Red Team (Adversarial) Testing?

Red Team testing is adversarial testing, like PEN testing and the solution is attacked.

Penetration testing focuses  on identifying and exploiting vulnerabilities within a system, while red team testing simulates a full-fledged cyberattack to assess an organization's overall security readiness.

Red team testing has its roots in AGILE mythology. Red team playing offensive, and Blue playing defensive.

Like PEN testing, it simulates real-world cyber-attacks, so both methods are complementary.

# What Can Red Team Testing Do?

**However, Red team testing lends itself to identifying a broader range of AI attacks. Why?**

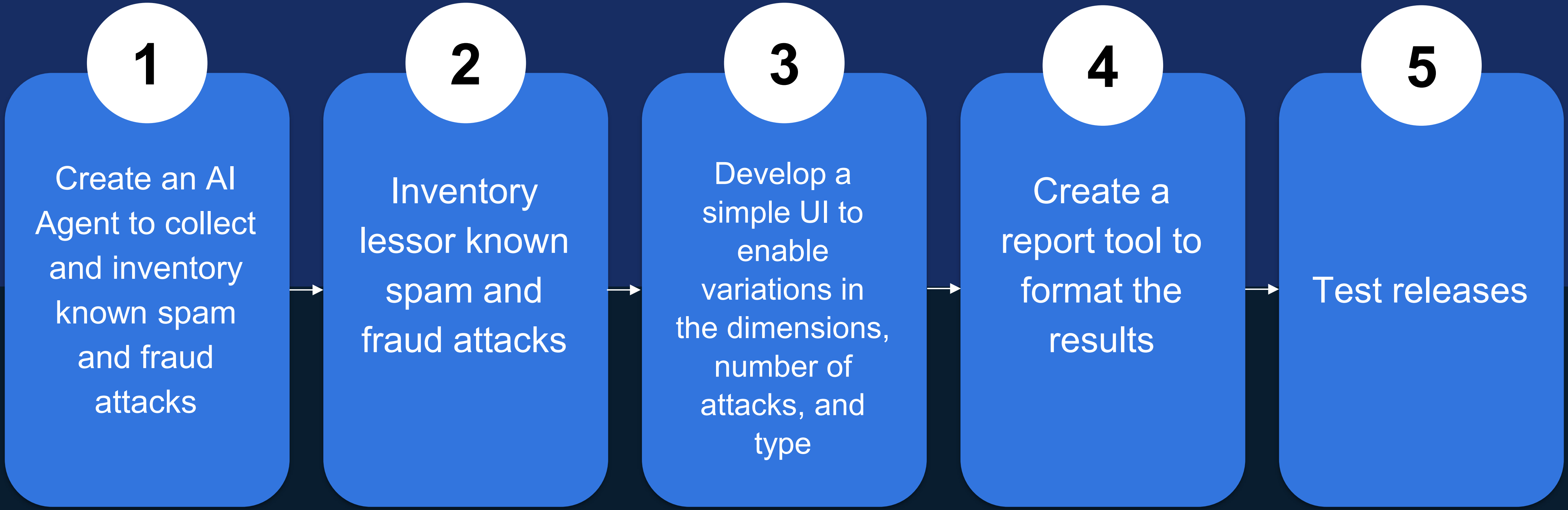
**Multi-dimensional attacks can be applied to a solution while in test**

**An attack based on AI tools can be modeled (1st dimension)**

**The attack can be adjusted for volume such as 10 TNs called, or 10,000 TNs (2nd dimension)**

**The attack can be adjusted for frequency such as snow shoeing or carpet bombing (3rd dimension)**

# An Approach to Red Team Testing



1. Create an AI Agent to collect and inventory known spam and fraud attacks

2. Inventory lessor known spam and fraud attacks

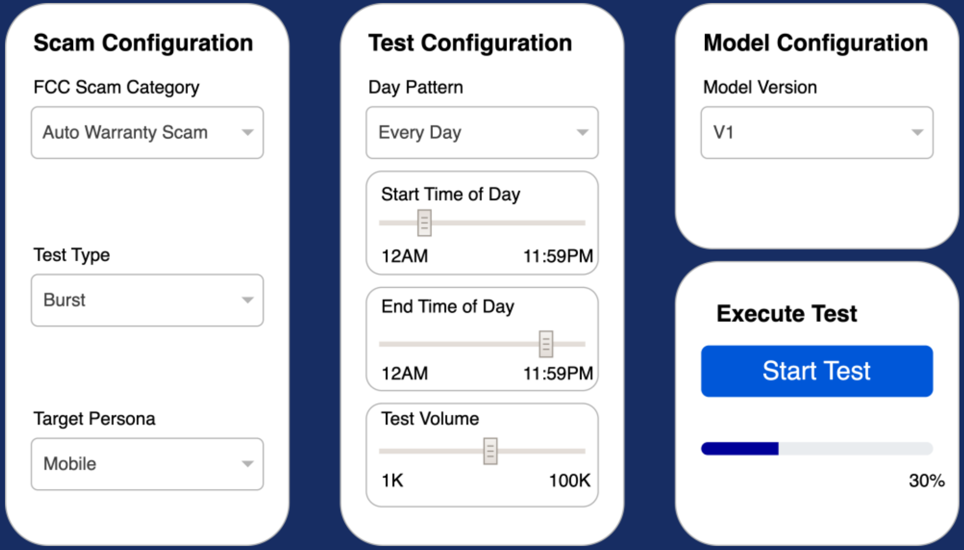
3. Develop a simple UI to enable variations in the dimensions, number of attacks, and type

4. Create a report tool to format the results

5. Test releases

# Sample of Red Team Testing Tool and Output

## Sample UI



**Scam Configuration**

FCC Scam Category: Auto Warranty Scam

Test Type: Burst

Target Persona: Mobile

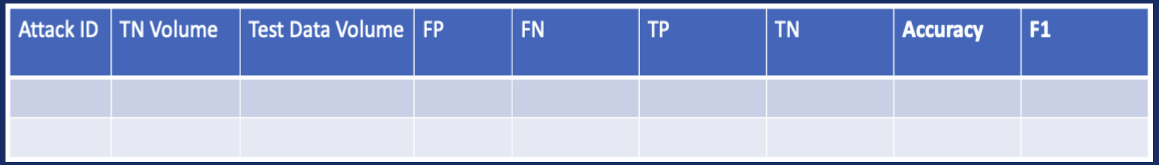
**Test Configuration**

Day Pattern: Every Day

**Model Configuration**

Model Version: V1

**Execute Test**



| Attack ID | TN Volume | Test Data Volume | FP | FN | TP | TN | Accuracy | F1 |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | | |
| | | | | | | | | |

## Test Report

- The **Burst Test** was performed on **80,000 TNS HoneyPot samples** collected on **June 19, 2025**, between **9 AM** and **5 PM** EST.
- Model **V1** successfully blocked **92%** of these samples.
  - Of the blocked samples:
    - **85%** were Auto Warranty Scams
    - **15%** were IRS Call Scams
    - **10%** were Utility Scams
- **8%** of samples were not successfully blocked by Model V1.
  - Of the unblocked samples:
    - **90%** were Auto Warranty Scams
    - **15%** were IRS Call Scams

# Why Red Team Testing?

**U.S. Regulators** suggested red team testing could be required for AI validation

Adversarial testing **validates** if a supplier implemented what they said or claimed they would

Combined with synthetic data, It can **stress test** prior to a new release or new market introduction

Adversarial testing can become an **addition** to or an **alternative** to **Honeypots** for performance validation

# Why Red Team Testing Verses a Honeypot?

Using a Honeypot, is like fishing with a line, bobber, and hook. Hoping something will come by and take the bait.

- Honeypots are telephone numbers with voice mail
- Honeypots can be dynamic such as Rapptor or TNS Sticky Honeypot

**But how often is it an *empty* hook?**

# Why Red Team Testing Verses a Honeypot ?

- Any spam or fraud attack which can be modeled, can then be tested.
    - Volume testing can be done from a single call to millions.
    - TN use can be single, multiple, rotated, or skipped randomly.
    - Any frequency can be modeled.

**In addition to the testing tool, the testing library then becomes a high value asset.**

Three dimensions of a single test

# Red Team Testing Opportunities

Regulated market – Can validate compliancy

The FCC lists known spam and fraud cases

**Is your system compliant?**

# Thank you!

Greg Bohl
gbohl@tnsi.com

tnsi.com