

TechDay India 2023



Ravi Prakash
Performance Engineering
Generative AI For Performance Engineering



“ Performance Engineering & Generative AI are the dynamic duo that optimize system performance , enhance user experience and drive business success in the ever-evolving digital world.”

Introduction



Today's Agenda

Welcome to the presentation on "Leveraging Generative AI for Enhanced Performance Engineering at Verizon".

- ❑ Why we Need Generative AI in Performance Engineering
- ❑ Solution - Leveraging Generative AI in Performance Engineering
- ❑ Implementation Approach and Proof of Concept(PoC)
- ❑ Conclusion - Empowering Performance Engineering with Generative AI

Need for Generative AI in Performance Engineering

Infrastructure Side:-

- ❑ **Specific Problem 1:** Manual calculation of pods and infrastructure resources leads to inaccuracies.
- ❑ **Impact:** Inefficient resource allocation, potential performance bottlenecks, and increased costs.
- ❑ **Challenges:** Complex workload patterns and dynamic traffic demand hinder accurate resource estimation.

Client Side:-

- ❑ **Specific Problem 2:** Manual calculation of virtual users (VUs), load, and TPS results in suboptimal test scenarios.
- ❑ **Impact:** Incomplete test coverage and potential performance regressions.
- ❑ **Challenges:** Difficulty predicting real-world user behavior and varying traffic patterns.



Solution - Leveraging Generative AI in Performance Engineering

Specific Solution to Problem 1

- ❑ Predict Future Demand
- ❑ Dynamic Resource Allocation
- ❑ Prevent Performance Bottlenecks

Techniques Used:

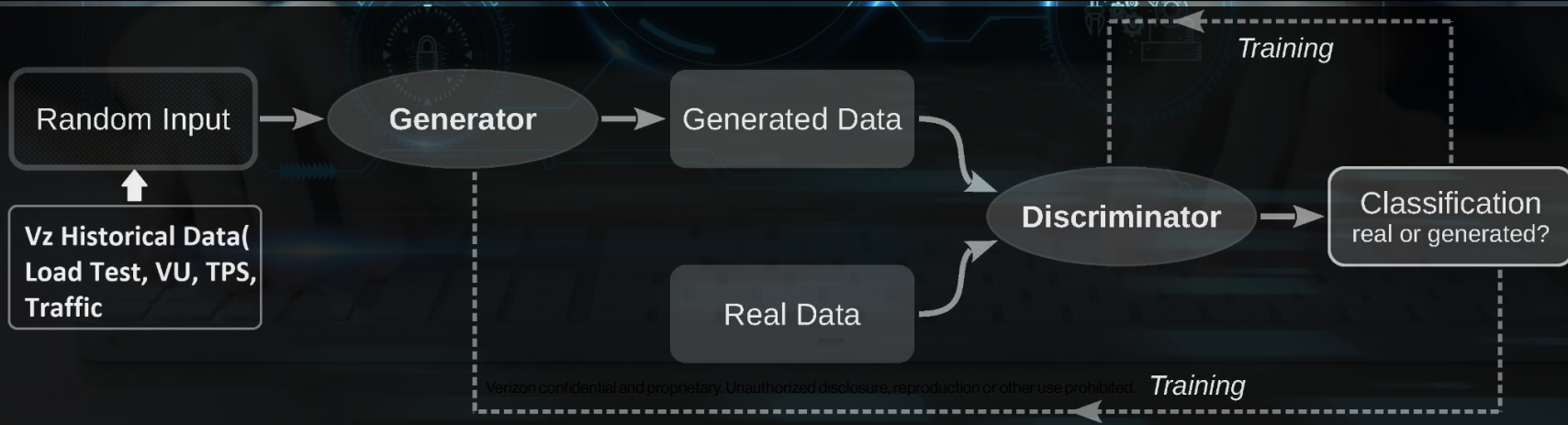
1. Generative adversarial networks (GANs)
2. Variational autoencoders (VAEs)
3. Natural language processing
4. Time Series Forecasting

Specific Solution to Problem 2:

- ❑ Generative Synthetic User Traffic
- ❑ Synthetic Load Generation
- ❑ Precise Test Scenarios

Techniques Used:

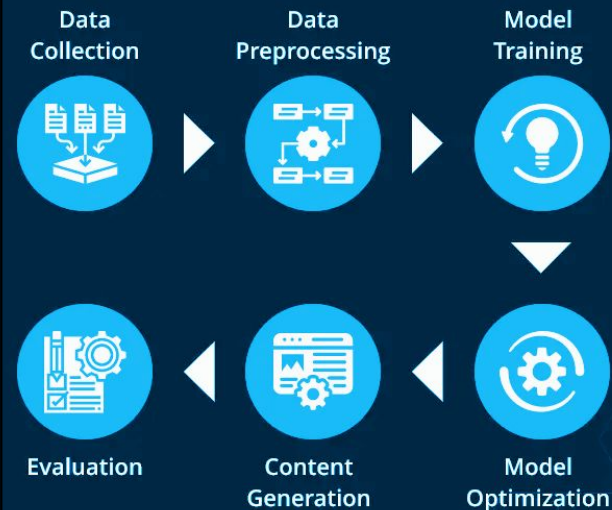
1. Reinforcement learning
2. Deep reinforcement learning
3. Multi-task learning



Implementation Approach and Proof of Concept(PoC)

Implementation for both Infrastructure & Client Side:

- ❑ **Data Collection:** Gather historical load test data and infrastructure metrics. (Kibana, NewRelic, Grafana etc)
- ❑ **Generative AI Model Training:** Train GANs or VAEs on data to learn workload patterns and correlations.
- ❑ **Real-time Resource Allocation:** Implement Generative AI for dynamic resource allocation.
- ❑ **Synthetic Load Generation:** Use Generative AI-generated synthetic VUs and workloads for testing.
- ❑ **PoC Evaluation:** Assess accuracy and efficiency of Generative AI-driven resource allocation.



Conclusion - Empowering Performance Engineering with Generative AI

Benefits of generative AI





verizon✓



**Thank
you**

