# Secure AI Model Sharing and Deployment

# Agenda

- GenAI in Platform Engineering
- Risks : GenAI Model Sharing
- Security Principles for GenAI Models
- Techniques for Secure GenAI Model Sharing
- Secure GenAI Deployment Strategies
- Utilizing Trusted Execution Environments
- Automated Security in GenAI Operations
- Case Study
- Future Trends
- Conclusion

# GenAI In Platform Engineering

GenAI revolutionizes platform engineering by automating complex tasks like code generation, system configuration, and error diagnosis, enabling personalization at scale, enhancing predictive analytics, and improving user interactions through advanced natural language processing capabilities.

GenAI not only inherits traditional AI vulnerabilities like bias, data privacy concerns, and challenges in explainability but also amplifies these issues due to its autonomous content generation, which can exacerbate biases and increase the risks of privacy breaches on a larger scale.

GenAI necessitates stringent security measures, including intellectual property safeguards, defenses against new adversarial attacks, and continuous system monitoring, alongside rigorous data protection and regular security audits.

# Risks : GenAI Model Sharing

**Data Poisoning Attack :**

- ***Risk*** **: Data poisoning subtly manipulates training data to corrupt AI models, potentially skewing outputs once deployed.**
- ***Detection*** **: Employ robust data validation and provenance checks to detect and correct malicious data alterations.**
- ***Mitigation*** **: Regularly retrain models with cleansed data to dilute the effects of any initial tampering, maintaining model integrity.**

# Model Inversion Attacks

- ***Risk*** : These attacks exploit model outputs to to infer sensitive details from the training data ,risking privacy breaches.
- ***Detection***: Implement differential privacy and output scrubbing to add randomness and restrict sensitive data leakage.
- ***Mitigation***: Enhance security with strict access controls and detailed auditing of who query the model and how it's used.

# Security Principles GenAI Models

- Zero-trust architecture is a security model based on the principle of "never trust, always verify."
- ***Continuous Verification*** : In GenAI, data inputs, model outputs, and user interactions occur at a rapid pace, often in real-time.  Zero-trust architecture mandates constant validation of these interactions to prevent unauthorized access and ensure the integrity of operations
- ***Minimized Insider Threats:*** GenAI systems are susceptible to risks from insiders who might misuse their permissions. Zero-trust mitigates this by limiting access to resources strictly to what is necessary for each user or service, based on their context and risk profile.
- ***Adaptability to Dynamic Environments:*** GenAI platforms often operate in dynamic, cloud-based environments where user devices and services are continuously changing. Zero-trust architectures thrive in such settings by dynamically adapting permissions and security measures based on fluctuating risk levels and contexts.

# Techniques for Secure GenAI Model Sharing

## Homomorphic Encryption :

- An encryption that allows computations to be carried out on ciphertext, generating an encrypted result which, when decrypted, matches the result of operations performed on the plaintext.
- enables the processing of sensitive data without exposing it. For example, a GenAI model could train on encrypted datasets, ensuring that the underlying data remains confidential throughout the computation process.
- Organizations can leverage HE to train GenAI models on encrypted data collected from multiple sources without accessing raw data directly. This is crucial in industries like healthcare or finance, where privacy is paramount.

## Secure API Strategies for GenAI Services ：

- Use robust authentication mechanisms like OAuth2 and OpenID Connect to verify the identity of users and services. Implement fine-grained authorization controls to ensure that users have access only to appropriate data and actions based on their roles.
- Protect APIs against abuse and Denial of Service (DoS) attacks by limiting the number of requests that users can make within a certain time frame
- Utilize TLS (Transport Layer Security) to encrypt data transmitted between clients and GenAI APIs. This prevents interception and tampering of sensitive data during transmission.
- Deploy API gateways to manage, monitor, and secure traffic to and from GenAI services. Gateways can provide additional layers of security such as IP whitelisting, threat detection, and logging for forensic analysis.

# Secure GenAI Deployment Strategies

***Digital Signature Verification***: Use digital signatures for model authenticity, implementing secure key management and version control to ensure only verified, unaltered GenAI models are deployed.

***Real-Time Monitoring***: Implement continuous monitoring with advanced tools to detect anomalies in GenAI behavior, utilizing sophisticated anomaly detection algorithms.

***Automated Response Protocols***: Integrate automated responses to swiftly address detected anomalies, enhancing security with feedback loops that refine detection and response strategies based on operational insights.

# Utilizing Trusted Execution Environments

***Isolated Execution:*** ***Trusted Execution Environments provide secure enclaves within processors, isolating sensitive data and model processing from the main operating system to prevent unauthorized access & tampering.***

***Enhanced Security Features:*** ***TEEs ensure data integrity and confidentiality through encryption and integrity checks, coupled with strict access controls.***

***Versatile Applications:*** ***Utilized in cloud and edge computing, TEEs are crucial for securing GenAI applications in sectors like healthcare and finance, where data security is paramount.***

# Automated Security in GenAI Operations

***Automated Vulnerability Scanners***: These tools identify, address vulnerabilities within AI models,data processes, integrating into CI/CD pipelines for early detection and continuous compliance.

***Behavioral Analytics and Code Analysis***: Tools that monitor AI system behaviors, analyze code to flag anomalies & known security risks, enhancing real-time security measures and developer awareness.

***Proactive Security and Compliance***: Continuous integration of security assessments ensures a proactive security posture, reduces security debt, and maintains compliance, fostering trust and reliability in GenAI operations.

# Case Study One : Healthcare

1. A healthcare company that needs to abide by HIPAA regulations can use GenAI system to improve diagnostic accuracy by training a model on vast set of anonymized patient image data.

2. The application can use homomorphic encryption to allow the GenAI model to process encrypted images directly, ensuring that data remained secure even during computation.

3. Digital signatures can be used to verify model integrity before each use, and TEEs can be used for secure model execution, isolating the model and data from the main hospital network.

# Future Trends

***Quantum-Resistant Encryption***: Development of post-quantum cryptography aims to secure systems against quantum computing threats, with standardization led by institutions like NIST.

***Predictive Security Models with GenAI***: Leveraging GenAI to analyze extensive data sets, predictive security models foresee vulnerabilities and potential threats, enabling proactive defense strategies.

***Integration and Standardization***: Emphasizes the need for integration of new cryptographic solutions into existing systems and standardization for widespread adoption.

# Conclusion

In today's evolving technology landscape, the role of security is important to not only protect against threats but also to maintain the trust and confidence of users and stakeholders.

As AI technologies keep advancing at a rapid pace, so will the sophistication of threats to against these systems.Its crucial the community-users,developers & regulators remain vigilant.

The complex landscape of AI and security demonstrates the necessity of a proactive approach—where security is not an afterthought, but a foundational component of all AI initiatives

# THANK YOU

linkedin.com/akshaysekar