A Report on Webinar
"The AI Revolution: Securing Our Digital Future"
Organized by Department of Computer Science & Engineering- Data Science
in association with ISTE, MITS
on 06.11.2024

Report Submitted by: Mrs. Roopa R, Assistant Professor, Department of Computer Science & Engineering- Data Science.
Resource Person Details: Mr. Noah Franklin, Regional Delivery Lead for AppSec America, Tech Mahindra, Bengaluru.
Venue: Scale up Room; Time: 10:00 AM to 12:00 PM
Mode of Conduct: Online
Report Received on 15.11.2024

Event Overview:

The Department of Computer Science & Engineering (Data Science), in association with ISTE, organized a webinar on "The AI revolution: Securing our digital future" The event aimed to provide students with insights into the transformative role of Artificial Intelligence in shaping the future of technology. The session focused on understanding AI's impact on various sectors, including cyber security, ethics, and the responsibilities associated with building a secure digital landscape.

Opening Remarks and Welcome Note:

The event began with a welcome address by Mrs. Roopa R, Assistant Professor from the Department of CSE (Data Science), who provided a brief overview of the event proceedings. She welcomed the resource person who graced the event and introduced him to the participants. The Head of the Department, CSE (Data Science) Dr. S. Kusuma addressed the gathering by explaining the importance of AI and responsibilities involved in creating a safe and resilient digital environment.

Key Note/ Resource Person Address:

The Resource Person of the event Mr. Noah Franklin delivered the keynote address. Mr. Noah completed his B.E in 2012. Presently he is working as Regional Delivery Lead for AppSec America, Tech Mahindra, Bengaluru. He has 12+ years of experience in web application security, network security and research. He is well versed in assessing information risk and facilitates remediation of identified vulnerabilities for IT security and IT risk across the enterprise. He is adept at developing team competency through training and knowledge sharing.

The Webinar highlighted with the following topics:

1. Recent AI-Based Cyber Data Leakage

The surge in AI-powered cyber-attacks has led to sophisticated data leakages, where malicious actors use AI to bypass traditional security measures. Techniques like deep fake impersonation and synthetic identity generation are commonly employed for unauthorized data access. Hackers now use AI to automate phishing and exploit machine learning models to analyze vulnerabilities in networks, amplifying the scale of potential breaches. AI-based tools can also evade anomaly detection, making it challenging to identify intrusions quickly. High-profile cases in finance and healthcare have demonstrated AI's potential in this sphere, highlighting the urgent need for robust countermeasures.

2. How E-mail Works

Email functions as a digital communication method using client-server technology. A sender composes a message and, upon hitting "send," the email client transmits it to an outgoing mail server via Simple Mail Transfer Protocol (SMTP). The server identifies the recipient's address, forwarding it across various servers until it reaches the receiver's incoming server, typically using IMAP or POP3. The recipient's email client then retrieves the email from the server, displaying it in the inbox. This process, usually completed within seconds, enables global communication with minimal delay.

3. Demo: AI-Powered Phishing and Detection

AI-powered phishing uses machine learning algorithms to mimic legitimate communication, generating realistic messages that lure users into disclosing personal information. Demonstrations show AI crafting emails with personalized details, often scraped from social media. However, AI-driven detection tools counter these threats by scanning message patterns, sender history, and analyzing linguistic cues. For example, models trained on phishing databases can identify unusual phrases or suspicious links, flagging emails as potentially harmful. This AI vs. AI battle highlights the evolving complexity of cyber defense.

4. Protecting Against AI-Related Data Leakage

To mitigate AI-driven data leakage risks, organizations are implementing strict access controls, encryption, and real-time monitoring of sensitive data. Anonymization techniques reduce the exposure of identifiable information in datasets, while differential privacy ensures data utility without revealing individual details. Employing secure model training and regular model audits are critical in detecting leaks. Additionally, adopting Zero Trust architectures helps prevent unauthorized access to data, ensuring that every access request is authenticated and authorized, even within internal networks**.**

5. Security Standards and Frameworks for LLM Model Processes

Securing Large Language Models (LLMs) involves adopting frameworks like NIST's AI Risk Management Framework, which emphasizes transparency, fairness, and accountability. ISO/IEC 27001 provides a baseline for information security management, applicable to AI systems managing sensitive data. Compliance with SOC 2 and GDPR ensures that data privacy measures are in place. Open AI and similar initiatives also provide AI-specific guidelines, advocating for model interpretability, secure deployment, and regular audits to minimize model exploitation risks, especially for models used in sensitive applications.

The session provides key outcomes for the students as follows:

- Students will identify methods by which AI is used in cyber-attacks and comprehend the implications of data leakage in various industries.
- Students will be able to describe the email process, including protocols like SMTP, POP3, and IMAP, and the flow of email from sender to receiver.
- Students will learn to recognize AI-driven phishing tactics and evaluate AI-based tools used in phishing detection, gaining hands-on experience in cyber defense.
- Students will explore techniques like encryption, anonymization, and Zero Trust principles to safeguard data against AI-driven leakage.
- Students will analyze security frameworks (NIST, ISO/IEC 27001) relevant to LLMs, understanding regulatory requirements and risk mitigation strategies for secure model deployment.

The session was concluded by Mrs. Roopa R, who delivered the vote of thanks. She thanked the resource person for delivering the 'Webinar' and the Head of the Department Dr. S. Kusuma, the Principal and the Management for giving the opportunity to initiate the event.